PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12P 19/34, C12N 15/10 — A1

(11) International Publication Number: **WO 96/38591**

(43) International Publication Date: 5 December 1996 (05.12.96)

(21) International Application Number: PCT/US96/08501

(22) International Filing Date: 3 June 1996 (03.06.96)

(30) Priority Data:

08/459,046	2 June 1995 (02.06.95)	US
08/462,355	5 June 1995 (05.06.95)	US
08/487,112	7 June 1995 (07.06.95)	US
60/006,809	15 November 1995 (15.11.95)	US
08/566,334	1 December 1995 (01.12.95)	US

(71) Applicant: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventor: GUEGLER, Karl, J.; 1048 Oakland Avenue, Menlo Park, CA 94025 (US).

(74) Agent: GLAISTER, Debra, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

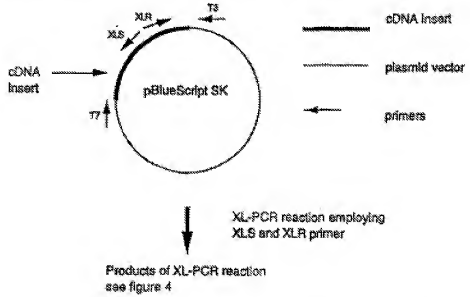
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IMPROVED METHOD FOR OBTAINING FULL-LENGTH cDNA SEQUENCES



(57) Abstract

A method for obtaining longer cDNA sequences is provided. The method utilizes a known genomic DNA sequence or a partial cDNA sequence, such as can be obtained from GenBank partial cDNAs. Two PCR primers are designed to correspond to the ends of the known partial sequence and to anneal to DNA in a cDNA library so as to initiate extension away from the known cDNA and the other primer. The primers are added to a cDNA library with appropriate enzymes and extend through additional DNA sequence to produce PCR products, which are subsequently purified and sequenced to provide new sequences. The new sequences are then compared with the known partial cDNA sequence for areas of overlap, and the sequence is extended beyond the overlapping areas to provide longer DNA sequence.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

IMPROVED METHOD FOR OBTAINING FULL-LENGTH cDNA SEQUENCES

TECHNICAL FIELD

The present invention is in the field of molecular biology and more particularly, in the field of recombinant DNA technology.

BACKGROUND ART

PCR has become a widely used nucleic acid amplification technique since it was first presented by Kary Mullis at the Cold Spring Harbor Symposium (Mullis K et al (1986) Cold Spring Harbor Symp Quant Biol 51: 263-273). PCR requires that a pair of primers be generated from known sequences. However, in many cases, sequence is available only from one end of a DNA segment. Several methods have been developed to sequence an entire gene once a partial nucleotide sequence is available. As more partial cDNA sequences become available in the world's genetic databanks, more efficient and economical methods will be sought for then obtaining the complete gene.

PCR has become a widely used technique to complete genes for which a partial sequence is already known. Gene-specific primers and primers located in the vector into which the cDNAs have been cloned are used for this purpose. However, this method is limited by the use of primers complementary to vector sequence which is common to all clones in the library. This results in an abundance of non-specific PCR-products which have to be cloned and sequenced. Multiple rounds of amplifications with nested primers might be required. These additional operations increase the incorporation of errors.

Gobinda, Turner and Bolander (1993) in PCR Methods and Applications 2:318-22 disclose "restriction-site PCR" as a direct method of retrieving unknown sequence which is adjacent to a known locus by using universal primers. First, genomic DNA is amplified in the presence of restriction site oligonucleotides and a primer

specific to the known region. Next, those products are subjected to a second round of PCR with the same restriction site oligonucleotides and another specific primer internal to the first one. Subsequently, the products of the last round of PCR are transcribed with an appropriate RNA polymerase and sequenced with a reverse transcriptase and an end-labeled specific primer internal to the second specific PCR primer. Gobinda et al. present data concerning Factor IX for which they identified a conserved stretch of 20 nucleotides in the 3' noncoding region of the gene.

Inverse PCR is the first method that reported successful acquisition of unknown sequences starting with primers based on a known region (Triglia T, Peterson MG, and Kemp DJ (1988) Nucleic Acids Res. 16:8186). Inverse PCR employs a strategy in which several restriction enzymes are used to generate a suitable fragment in the known region. The segment is then circularized by intramolecular ligation and used as a PCR template with divergent primers created from the known region. However, the requirement of multiple restriction enzyme digestions followed by multiple ligations (even before PCR is started) make the procedure slow and expensive (Gobinda et al. *Supra*).

Capture PCR, first disclosed by Lagerstrom M, Parik J, Malmgren H, Stewart J, Patterson U and Landegren U (1991) PCR Methods Applic. 1:111-19, is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. As noted by Gobinda et al. *supra*, that method also requires multiple restriction enzyme digestions and ligation of an engineered double-stranded primer before PCR. Although the restriction and ligation reactions are carried out simultaneously in this method, the requirement of extension reaction, immobilization of the extended product, two rounds of PCR and purification of template prior to sequencing render it cumbersome and time consuming as well.

Walking PCR, disclosed by Parker JD, Rabinovitch PS, and Burmer GC (1991) Nucleic Acids Res 19:3055-60, teaches a method for targeted gene walking via PCR. Although this method also permits retrieval of unknown sequence, Gobinda et al, *supra*, note that it requires oligomer-extension assay followed by identification and gel purification of the desired band prior to sequencing. Such extra steps again limit the applicability of the method.

The enzymes originally used in PCR were limited in their ability to reliably amplify long pieces of nucleic acids over 3kb. One of the explanations for this limitation seems to be the misincorporation of nucleotides resulting in non-basepairing mismatches which these enzymes often fail to extend.

Only the mixture of two enzymes, rTth DNA-Polymerase and Vent, the latter of which has so-called "proofreading" activity, and the optimization of amplification conditions finally overcame this limitation and made amplification of pieces of DNA of up to 40kb possible.

The most common way to identify genes expressed in a certain tissue at a certain time is the isolation of the mRNA of that particular tissue and the conversion of this mRNA into so-called cDNA (complementary DNA). This cDNAs are subsequently cloned into a vector (plasmid or Lambda) and amplified by transfection into E.coli cells resulting in a so-called cDNA library.

First and most important to researchers attempting to obtain a complete gene is that the enzymes used in converting mRNA into cDNA are limited in their ability to produce complete copies of the existing mRNAs. This requires the researcher to isolate multiple cDNA clones of the gene of interest using specific probes and analyze each of these isolates for a complete cDNA of the gene of interest. This process is called screening of cDNA libraries.

A major problem facing molecular biologists is finding the most efficient method to use to obtain a full-length cDNA from a

partial sequence. Such sequences are appearing with increasing frequency in GenBank, from commercial cDNA libraries and privately prepared libraries. The inventive method disclosed herein is a contribution to that art.

DISCLOSURE OF THE INVENTION

An improved method for extending the DNA sequence of a known fragment of DNA sequence is provided. The method may be used for extending known DNA sequences of genomic or cDNA origin. The method utilizes the polymerase chain reaction (PCR) and includes the steps of:

a) combining a first and second PCR primer with nucleic acid from a cDNA library, or pools of cDNA libraries, expected to contain said partial cDNA, or said partial cDNA that has been extended, or a genomic library, under conditions suitable for synthesis of nucleic acid PCR products from the first and second primers, wherein said first and second primers are capable of annealing to opposite strands of the partial cDNA or genomic DNA and initiating nucleic acid synthesis in an outward manner and wherein the first primer is capable of being extended by DNA polymerase in an antisense direction and the second primer is capable of being extended in a sense direction,

b) purifying the PCR products, and

c) identifying extended nucleotide sequences derived from said partial cDNA or said genomic DNA. In one embodiment of the present invention, the method of identifying the extended nucleotide sequences comprises nucleic acid sequencing. In another embodiment of the present invention, the method proceeds with repeating steps 6a through 6c on the nucleotide sequences identified in step 6c.

In another embodiment of the present invention, there is a method for extending the nucleotide sequence of a partial complementary DNA (cDNA) using polymerase chain reaction (PCR), comprising the steps of a) combining a first and second PCR primer

with nucleic acid from a cDNA library, or pools of cDNA libraries, expected to contain said partial cDNA, or said partial cDNA that has been extended, or a genomic DNA library, under conditions suitable for synthesis of nucleic acid PCR products from the first and second primers, wherein said first and second primers are capable of annealing to opposite strands of the partial cDNA and initiating nucleic acid synthesis in an outward manner and wherein the first primer is capable of being extended by DNA polymerase in an antisense direction and the second primer is capable of being extended in a sense direction,

b) purifying the PCR products,

c) ligating the purified PCR products under conditions suitable for the formation of circular, closed nucleic acid,

d) transforming a host cell with the circular, closed nucleic acid and culturing the transformed host cell under conditions suitable for growth,

e) recovering said circular closed nucleic acid from the cultured, transformed host cell, and

f) identifying extended nucleotide sequences derived from said partial cDNA or said genomic DNA.

The present invention also provides a method for extending known genomic DNA sequences which may be used for the detection and amplification of 5' untranslated nucleotide sequences and/or promoter sequences.

Also provided is an isolated DNA molecule comprising SEQ ID NO:11, the DNA for a novel human purinergic P2U receptor.

Also provided is an isolated DNA molecule comprising SEQ ID NO:12, the DNA for a novel human C5a-like seven transmembrane receptor.

These and other objects, advantages and features of the present invention will become apparent to those persons skilled in the art upon reading the details of the structure, synthesis, formulation and usage as more fully set forth below, reference

being made to the accompanying figures forming a part hereof.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a flow chart of the steps in the inventive method.

Figure 2 shows a typical plasmid obtained from the excision process of a lambdaZAP cDNA library. Typically 250-300 base pairs of the sequence are obtained in the high-throughput sequence operation. The clone is partially sequenced from the 5' end with T3 as a sequencing primer.

Figure 3 is a representation of the next step, in which pBLUESCRIPT SK plasmids in a cDNA library are used as a template and the two specially designed primers (XLR and XLS) amplify plasmids containing the gene of interest. Only plasmids containing priming sites for both XL-PCR primers and the gene of interest will be amplified during the XL-PCR reaction.

Figure 4 is a representation of the amplified DNA segments which have been obtained through the XL-PCR reaction and consequently purified after separating the products on an agarose gel. For best results, the cDNA library used as a template should be synthesized by random priming to assure the availability in this step of different amplified length of DNA (3' end) between the XLS priming site and the T7 priming site in the vector. The length of the 5' end (between the XLR priming site and the T3 priming site) in the vector will vary in size depending on how much of the mRNA of the gene of interest had been converted into cDNA during the cDNA library synthesis.

Figure 5 shows how the purified DNA segments containing the plasmid and the gene of interest are religated to form a circular plasmid and transformed into bacteria for amplification. Here chemically competent E. coli cells were transformed and grown on petri dishes containing LB agar and 25 mg/L carbenicillin (2XCarb) for antibiotic selection.

Figure 6 shows schematically how pure samples of clones were

obtained from the different E. coli colonies grown in the procedure shown in Figure 5 (also Step 1 purification, Step 2 religation and Step 3 transformation in Figure 6). These clones are screened in Step 4 for additional sequence of the gene of interest at the 5' end. For this purpose the clones were analyzed by a PCR reaction employing the XLR primer and the T3 vector primer. The size of the resulting product will indicate how much additional sequence upstream of the XLR priming site each clone contains.

Figures 7A through 7H show the results of the inventive method, in which a partial sequence from Incyte clone 14770, which was similar to heat shock protein 90, was successively sequenced to obtain a full-length cDNA.

Figures 8A through 8F show the results of the inventive method, in which a partial sequence from Incyte clone 87058 which was similar to cathepsin was successively sequenced to obtain extensions of the cDNA.

MODES FOR CARRYING OUT THE INVENTION

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents and publications referred to herein are incorporated by reference herein.

Before the present compounds, variants, formulations and methods for making and using such are described, it is to be understood that this invention is not limited to the particular compounds, variants, formulations or methods described, as such enzymes, formulations and methodologies may, of course, vary. The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting since the scope of protection will be limited only by the appended claims.

In the specification and appended claims, the singular forms

"a", "an" and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a high-fidelity PCR enzyme" includes mixtures of such enzymes and any other enzymes fitting the stated criteria, reference to the method includes reference to one or more methods for obtaining full-length cDNA sequences which will be known to those skilled in the art or will become known to them upon reading this specification.

The present method provides a way to utilize a genomic DNA library or a plasmid cDNA library (either obtained by cloning cDNAs directly into a plasmid vector or by converting a Lambda library into a plasmid library by known methods e.g. Lambda ZAP excision or Lambda &IPLOCK conversion) which has been used for sequencing cDNAs, as a source to obtain much longer DNAs and in certain cases complete genes of partially known DNA sequences. The steps disclosed herein are based on cDNA libraries but equally apply to genomic DNA libraries.

This new method utilizes PCR kits which enable the researcher to amplify long pieces of DNA. The XL-PCR amplification kit (Perkin-Elmer) was employed. However, equivalent products may be available from other major suppliers. This novel method allows one person to process multiple genes (up to 96 genes) at a time and obtain extended or complete sequence (possibly full-length) of the cDNAs of interest within 6-10 days. This compares very favorably with current competitive methods like screening with labelled probes which allow one worker to process only about 3-5 genes and obtain initial results in 14-40 days. This represents an increase in throughput of at least 1000%.

This increased efficiency is possible because of the inventive combination of steps shown in the flow chart (Figure 1). First, primer design and synthesis (based on a known partial sequence) can be performed in about two days. The PCR amplification can be performed in 6-8 hours. Multiple libraries

can be pooled and therefore screened at the same time. The next steps of purification and ligation take about one day. Then transformation and growing up the bacteria take one day. Then screening for clones with additional sequence of the genes of interest by PCR takes approximately five hours. The next steps of DNA preparation and sequencing of the selected clones can be performed in about one day. This totals 6-7 days. At the end of this time, one has usually obtained a much longer cDNA sequence, assuming such a longer cDNA existed in the libraries than what was initially sequenced. If the new sequence is a complete gene, then the goal has been reached. If the complete sequence has not been obtained, one still has a much longer sequence than before, and this longer sequence can be used to design primers to repeat the procedure on the same or another library. The choice of library is up to the researcher, but a preferred library is one that has been size-selected to include only larger cDNAs.

This method presumes that one already has partial cDNA sequences, either from a publicly available database or the scientist's own earlier research, including but not limited to earlier preparation of a cDNA library whose cDNAs have been partially sequenced. The cDNA library may have been prepared with oligo dT or random primers. The difference between oligo dT and randomly primed libraries is that a randomly primed library will have more sequences which contain 5' ends of cDNAs. A randomly primed library may be particularly useful for further work when the oligo dT library does not yield a complete gene. Random priming of the library also helps yield more cDNA sequences of different lengths. Library preparation techniques which promote longer insert sizes will in turn permit the sequencing of more complete cDNAs. Obviously, the larger the protein, the less likely it is that the complete cDNA will be found in a single plasmid.

Figure 2 shows a typical plasmid containing a cDNA which had

been partially sequenced from the 5' end with T3 as a primer. The top darkened portion represents the insert containing the gene of interest.

Step 1: PCR-amplification of cDNA-clones containing the gene of interest

The first step of this method requires the design of two primers based on the known sequence. The known sequence can be obtained by those skilled in the art either by a wet lab method or from the many publicly available DNA databases. One primer is synthesized to be extended in an antisense direction (XLR) and the other in the sense direction (XLS or XLF). In effect, the primers are designed to anneal to either end of the known sequence and to be extended "outward" from there to generate amplicons containing new, unknown sequences of the genes of interest. This is different from typical PCR, in which the primers are designed to amplify a known sequence in a direction "inward" toward each other.

The primers need to be designed in a way displaying optimal criteria for extra long PCR. A program like Oligo 4.0s (National Biosciences, Inc., Plymouth MN) can be employed for this purpose. In general primers should be 22-30 nucleotides in length, consist of a GC content of 50% or more and anneal at 68°C-72°C to the target. Hairpin structures and primer-primer dimerizations must be avoided.

Primers varying from the conditions described above may result in amplification of the desired targets providing extension conditions have been adjusted.

Figure 3 shows the next step, in which a cDNA library is used as a template and the two primers (XLR and XLS) amplify plasmids containing the gene of interest. In this step, it is very helpful to use PCR enzymes which provide high fidelity and copy long sequences, such as that provided in the XL-PCR kit (Part No. N808-0182, Perkin Elmer, Applied Biosystems, Foster City, CA).

Generally, kit instructions should be followed, including suggestions to optimize concentrations of various reagents. In the examples disclosed *infra*, 25pMol of each primer worked well. Template (plasmid library) concentrations can be varied (see Examples *infra* for details). It is essential to thoroughly resuspend the enzyme in solution prior to use, especially if the solution has been stored at -20°C. If the enzyme is not adequately resuspended, its effectiveness is impaired. The preferred system is setup initially in two layers, employing Ampliwax® PCR Gems. However, efficiency can be increased by avoiding the use of these Gems and initiating amplification by using the "hot-start" technique by adding Magnesium, which is essential for amplification, at 82° C.

Although various cycling conditions are detailed in the examples *infra* , the following cycling conditions have been found to be optimal with the MJ PCT200 thermocycler (MJ Research, Watertown, MA). Times and temperatures may be varied to optimize conditions in different thermocyclers.

Step 1 94° for 60 sec (initial denaturation)
Step 2 94° for 15 sec
Step 3 65° for 1 min
Step 4 68° for 7 min
Step 5 Repeat step 2-4 for 15 additional times
Step 6 94° for 15 sec
Step 7 65° for 1 min
Step 8 68° for 7 min + 15 sec/cycle
Step 9 Repeat step 6-8 for 11 additional times
Step 10 72° for 8 min
Step 11 4° for 0.00 sec (to hold at 4°)

At the end of these 28 cycles, 50 µl of the reaction mix is removed; on the remaining reaction mix, an additional 10 additional cycles are run, as outlined below:

Step 1 94° for 15 sec
Step 2 65° for 1 min
Step 3 68° for (10 min + 15 sec)/cycle
Step 4 Repeat step 1-3 for 9 additional times
Step 5 72° for 10 min

Next a 5-10 µl aliquot of the reaction mixture can be analyzed on a mini-gel to determine which reactions were successful.

Step 2: Purification of amplicons containing the gene of interest

Figure 4 is a graphical representation of the amplified cDNA segments which have been separated on an agarose gel. Note that there are a variety of lengths of cDNA. Although the rest of the method could be performed using all extended cDNA species, the method can proceed optionally after selecting the largest products (likeliest to provide the remainder of the full-length gene). Some of the larger species may in fact be hybrid clones which contain two cDNA inserts as a result of malfunction during the cDNA library construction which may represent an incomplete digestion with the restriction enzyme at the end of the cDNA synthesis. Such amplified hybrid clones, also called chimera, could result in overlooking the correct targeted extensions.

Successful reaction products should be purified on an agarose gel (preferentally low agarose concentrations 0.6-0.8% should be used) or other appropriate method. An appropriate volume of reaction mixture should be loaded to obtain good separation of the products and to separate them from the plasmid library (template) still in the reaction mixture. Contamination with the template cDNA library will result in transformants which don't contain the desired gene and will require an extensive screening of many colonies. The bands representing the genes of interest are then cut out of the gel and purified using a method like the QIAQuick gel extraction kit (Qiagen, Inc., Chatsworth, CA).

Step 3: Cloning of amplicons containing the gene of interest

Eventual overhangs are converted into blunt ends to facilitate religation and cloning of the products. For this purpose, Klenow enzyme (3 units/reaction mixture) and dNTP's (0.2 mM final concentration) are added and the reaction is incubated at room temperature for 30 min. The Klenow enzyme is then

inactivated by incubating the reaction at 75° for 15 min.

The products are then ethanol precipitated and redissolved in 13 µl of ligation buffer containing 1 mM ATP. 1ml T4-DNA ligase (15 units) and T4 Polynucleotide kinase (5 units) are added and the reaction is incubated at room temperature for 2-3 hours or overnight at 16°C.

3µl of the ligation mixture are transformed into 40ml of competent E.coli cells (prepared with a standard protocol). 80µl of SOC medium are added and after 1 hour of recovery of the cells at 37°C the whole transformation mixture is plated on LB-agar 2XCarb-containing petri plates.

Step 4: Screening of cloned products

The next day 8 or 12 colonies are randomly picked from each plate and grown in individual wells of a sterile 96-well microtiter plate (e.g. 96 Well Cell Culture Cluster, Catalog No. 3799, Costar Corp., Cambridge, MA 02140). Each well contains 150ml of LB/2XCarb medium. Thus, each row of the microtiter plate contains twelve clones from the same extension reaction. The cells are grown over night at 37°C.

The next day, 5 µl of these overnight cultures are tranferred into a non-sterile 96-well plate (Falcon 3911 Microtest III™, Flexible Assay Plate, Becton Dickinson, Oxnard, CA) and diluted 1:10 with water. 5µl of each dilution are then transferred into a PCR array (e.g., Cycleplate, Robbins Scientific Corp., Sunnyvale, CA). To obtain a 1X final concentration of PCR reagents, 15 µl of a 1.33X concentrated PCR mix are added to each well. Another way of efficient screening for extension products is the multiplex PCR method where multiple specific primers are pooled and submitted to the same reaction, therefore increasing the efficiency of setting up the screening mixtures. Addition of the PCR-template (individual cultures) has been improved by the use of a 96-pin tool with which an aliquot of all 96 cultures grown as described

above can be transferred into the PCR-screening mix in a matter of 1-2 minutes.

For PCR amplification, the final concentrations are 1X for PCR mix, 5 μM of each of a vector primer and one or both of the gene specific primers used for the original extension reaction and 0.75 units of Taq polymerase are added to each well. Amplification generally was performed using the following conditions:

Step 1 94°C for 60sec
Step 2 94°C for 20sec
Step 3 55°C for 30sec
Step 4 72°C for 90sec
Step 5 repeat steps 2-4 for an additional 29 times
Step 6 72°C for 180sec
Step 7 4°C for ever

Aliquots of these PCR reactions are run on agarose gels together with molecular weight markers. The size of the resulting PCR products will allow direct determination of how much additional sequence the selected clones contain compared to the original partial cDNA. The efficiency of the method has been further improved by using the resulting PCR-products directly for sequencing thus avoiding the necessity of preparing plasmids.

The appropriate clones are selected and grown for plasmid preparation and sequencing.

Plasmid preparations are made with standard kits familiar to those skilled in the art. Examples include the PROMEGA Magic MINIPREP and the AGTC alkaline lysis kit.

Sequencing is performed employing standard automated ABI sequencing equipment and protocols using either dye-primer or dye-terminator kits.

Sequence processing and assemblage of the sequencing data are performed using standard ABI software, including INHERIT™ analysis and the Power assembler.

INDUSTRIAL APPLICABILITY

Example 1

For the initial method evaluation, a known gene was selected. A partial sequence of the human 90-kDa heat-shock protein gene (HUMHSP90, accession M16660) had been identified in a THP-1 library. This partial sequence (Incyte clone T-014201) initiated at base 1127 of the sequence with accession number M16660.

1.1 Primer design

Two primers were designed to perform the method described in the invention.

Primer 1 (XLR) 5' AGC TGT CCA TGA TGA ACA CAC G 3'
(1180-1159)

Primer 2 (XLS) 5' AAT AGG CAC CAC ACC AAC TGA G 3'
(2011-2032)

1.2 Template preparation

A THP-1 cDNA library constructed into the LambdaZAP vector (Stratagene) was converted into a plasmid library following the mass excision protocol. Plasmids of the excised libraries were prepared using the Quiagen Midi plasmid purification kit.

1.3 XL-PCR reaction set-up

The extension reactions were prepared following the instructions provided with the GeneAmp XL PCR Kit (Part No. N808-0182) from Perkin Elmer. A two layer system was set up as follows:

The lower reagent mix was prepared by pipetting the following components into a 0.2ml MicroAmp reaction tube.

Lower reagent mix preparation:

Water		13.6 µl
3.3X buffer		12.0 µl
dATP	(10mM)	2.0 µl
dCTP	(10mM)	2.0 µl

dGTP	(10mM)	2.0 µl
dTTP	(10mM)	2.0 µl
Primer XLS	(50µM)	1.0 µl
Primer XLR	(50µM)	1.0 µl
Mg(OAc)2	(25mM)	4.4 µl
Total lower reagent mix		40.0 µl

One AmpliWax™ gem was added to the tube. The wax was melted by incubating the reaction tubes at 75°C for 5 minutes. Then the tubes were cooled down to 4°C.

Upper reagent mix preparation:

3.3X buffer	18.0 ml
rTth DNA Polymerase	2.0 ml
Total upper enzyme mix	20.0 µl

20 µl of the enzyme/buffer mix are added to each tube and kept separated from the lower mix by the wax layer.

Addition of template:

The template DNA (excised library) was diluted to an appropriate concentration in water and then added to the upper mix. Mixing of the components is not necessary.

Template (6.25ng/ml)	40.0 µl
Final volume	100.0 µl

1.4 XL-PCR amplification

For amplification the following protocol was employed:

Step 1 94° for 60 sec (initial denaturation)
Step 2 94° for 15 sec
Step 3 65° for 1 min
Step 4 68° for 7 min
Step 5 Repeat step 2-4 for 15 additional times
Step 6 94° for 15 sec
Step 7 65° for 1 min
Step 8 68° for 7 min + 15 sec/cycle
Step 9 Repeat step 6-8 for 11 additional times
Step 10 72° for 8 min
Step 11 4° for 0.00 sec (to hold at 4°)

1.5 Purification of amplified products

30 µl of the amplified products were run on a 0.7% agarose gel for 16 hours. Visible DNA bands were then cut out and purified using the QIAquick gel purification kit.

1.6 Cloning of amplified products

Klenow enzyme (3 units/reaction) and dNTP's (0.2mM final concentration) were added and the reactions were incubated at room temperature for 30 min followed by incubation at 75° C for 15 min. The products were then ethanol precipitated and redissolved in 13 µl of ligation buffer containing 1mM ATP. T4-DNA ligase (15 units) and T4 Polynucleotide kinase (5 units) were added, and the reaction was incubated at room temperature for 3 hours.

3µl of the ligation mixture were transformed into 40 ml of competent E.coli cells. After heatshocking the cells at 42° C for 45 seconds, 80 µl of SOC medium were added, and the cells were allowed to recover at 37° C for 1 hour. The whole transformation mixture then was plated on LB-agar/2XCarb-containing petri dish plates.

1.7 Screening of cloned products

The next day 10 colonies were randomly picked and grown

overnight in Falcon 2059 tubes (Becton Dickinson, Oxnard, CA) containing 3 ml of LB-broth with 2X Carb.

5 µl of the cultures were diluted 1:10 with water and 5 ml of this dilution were transferred into MicroAmp™ PCR tubes (Perkin Elmer, Applied Biosystems, Foster City, CA).

15 µl of a 1.33X concentrated PCR mix were added to each well.

The 1.33 x concentrated PCR mix contained the following components:

10X PCR-buffer	2.0 µl
2mM dNTPs	2.0 µl
M13 rev primer (0.01mM)	1.0 µl
Primer 2 (XLR, 0.01mM)	1.0 µl
Taq Polymerase	0.15 µl
Water	8.85 µl
Final Volume	15.0 µl

The PCR cycling conditions were choosen as follows:

Step 1 94° C for 60sec
Step 2 94° C for 20sec
Step 3 55° C for 30sec
Step 4 72° C for 90sec
Step 5 repeat steps 2-4 for an additional 29 times
Step 6 72° C for 180 sec
Step 7 4° C for ever

Aliquots of the amplified products were run on a 0.8% agarose gel in parallel with the 1 kb DNA ladder (Life Technologies, Gaithersburg, MD 20897). Appropriate plasmids containing different size inserts were selected for sequencing analysis.

1.8 Sequencing analyis of cloned products

The DNA of the selected clones was prepared using the

WizardTM Minipreps DNA Purification System (Promega Corporation, Madison, WI) following the instructions of the manufacturer. Sequencing reactions were performed using the PRISMTM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Part No 401628, Perkin Elmer, Applied Biosystems, Foster City, CA).

1.9 Analysis of sequenced products

Three clones were selected for sequencing (14201.3, 14201.5, 14201.13). The sequences obtained (SEQ ID NOS:3-5, respectively) were aligned using the DNASIS Multiple sequence alignment program. Clone 14201.3 initiated at base 24 of the published sequence (HUMHSP90), clone 14201.5 initiated at base 13 of the published sequence and clone 14201.13 initiated at base 538 of the published sequence, the original clone (14201) initiated at base 1127 of the published sequence.

Figure 7A-7H shows an alignment of the obtained sequences with the published human Hsp 90 nucleotide sequence. Clones 14201.3 and 14201.5 contain part of the 5' untranslated region and therefore the full coding region of the gene has been obtained.

Example 2

For further method evaluation, a second known gene was selected. A partial sequence from a liver library was found to be related to that of the human cathepsin B gene (accession L16510, HUMCATHB, SEQ ID NO:6). This partial sequence (Incyte clone 87058, SEQ ID NO:7) initiated at base 1066 of the sequence with accession number L16510.

2.1 Primer design

Two primers were designed to perform the method described in the invention:

Primer 1 (XLR) 5' AAG CCA TTG TCA CCC CAC TCA G 3'
(1103-1082)

Primer 2 (XLS) 5' GGT TCA CTG TGG AAT CGA ATC 3'
(1125-1145)

2.2 Template preparation

A liver cDNA library constructed into the LambdaZAP vector (Stratagene) was converted into a plasmid library following the mass excision protocol. Plasmids of the excised libraries were prepared using the Quiagen Midi plasmid purification kit.

2.3 XL-PCR reaction set-up

The extension reactions were prepared following the instructions provided with the GeneAmp XL PCR Kit (Part No. N808-0182) from Perkin Elmer. A two layer system was set up as described below. The lower reagent mix was prepared by pipetting the following components into a 0.2ml MicroAmp reaction tube.

Lower reagent mix preparation:

Water		13.6 µl
3.3 x buffer		12.0 µl
dATP	(10mM)	2.0 µl
dCTP	(10mM)	2.0 µl
dGTP	(10mM)	2.0 µl
dTTP	(10mM)	2.0 µl
Primer XLS	(50µM)	1.0 µl
Primer XLR	(50µM)	1.0 µl
Mg(OAc)2	(25µM)	4.4 µl
Total lower reagent mix		40.0 µl

One AmpliWax® gem was added to the tube. This was melted by incubating the reaction tubes at 75°C for 5 minutes. Then the tubes were cooled down to 4°C.

Upper reagent mix preparation:

3.3X buffer	18.0 µl
rTth DNA Polymerase	2.0 µl

Total upper enzyme mix	20.0 µl

20 µl of the enzyme/buffer mix were added to each tube and kept separated from the lower mix by the wax layer.

Addition of template:

The template DNA (excised library) was diluted to an appropriate concentration in water and then added to the upper mix. Mixing of the components is not necessary.

Template (6.25ng/µl)	40.0 µl
Final volume	100.0 µl

2.4 XL-PCR amplification

For amplification the following protocol was employed:

Step 1	94° for 60 sec (initial denaturation)
Step 2	94° for 15 sec
Step 3	65° for 1 min
Step 4	68° for 7 min
Step 5	Repeat step 2-4 for 15 additional times
Step 6	94° for 15 sec
Step 7	65° for 1 min
Step 8	68° for 7 min + 15 sec/cycle
Step 9	Repeat step 6-8 for 11 additional times
Step 10	72° for 8 min
Step 11	4° for 0.00 sec (to hold at 4°)

2.5 Purification of amplified products

30 µl of the amplified products were run on a 0.7% agarose gel for 16 hours. Visible DNA bands were then cut out and purified using the QIAQuick gel purification kit.

2.6 Cloning of amplified products

Klenow enzyme (3 units/reaction) and dNTP's (0.2mM final concentration) were added, and the reactions were incubated at room temperature for 30 min followed by incubation at 75°C for 15

min.

The products were then ethanol precipitated and redissolved in 13 µl of ligation buffer containing 1mM ATP. T4-DNA ligase (15 units) and T4 Polynucleotide kinase (5 units) were added, and the reaction was incubated at room temperature for 3 hours.

3 µl of the ligation mixture were transformed into 40 µl of competent E.coli cells. After heatshocking the cells at 42°C for 45 seconds, 80 µl of SOC medium were added; and the cells were allowed to recover at 37o C for 1 hour. The whole transformation mixture then was plated on LB-agar 2x Carb-containing petri dishes.

2.7 Screening of cloned products

The next day 10 colonies were randomly picked and grown overnight in Falcon 2059 tubes (Becton Dickinson, Oxnard, CA 93030) containing 3 ml of LB-broth with 2X Carb.

5 µl of the cultures were diluted 1:10 with water and 5 µl of this dilution were transferred into MicroAmpTM PCR tubes (Perkin Elmer, Applied Biosystems, Foster City, CA).

15 µl of a 1.33 x concentrated PCR mix were added to each tube.

The 1.33 x concentrated PCR mix contained the following components:

10 x PCR-buffer	2.0 µl
2mM dNTPs	2.0 µl
M13 rev primer (0.01mM)	1.0 µl
Primer 2 (XLR, 0.01mM)	1.0 µl
Taq Polymerase	0.15 µl
water	8.85 µl
Final Volume	15.0 µl

The PCR cycling conditions were as follows:

Step 1 94°C for 60sec
Step 2 94°C for 20sec
Step 3 55°C for 30sec
Step 4 72°C for 90sec
Step 5 repeat steps 2-4 for an additional 29 times
Step 6 72°C for 180sec
Step 7 4°C for ever

Aliquots of the amplified products were run on a 0.8% agarose gel in parallel with the 1kb DNA ladder (Life Technologies, Gaithersburg, MD 20897). Appropriate clones containing different size inserts were selected for sequencing analysis.

2.8 Sequencing analysis of cloned products

The DNA of the selected clones was prepared using the WizardTM Minipreps DNA Purification System (Promega Corporation, Madison, WI) following the instructions of the manufacturer. Sequencing reactions were performed using the PRISMTM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Part No 401628, Perkin Elmer, Applied Biosystems, Foster City, CA).

2.9 Analysis of sequenced products

Three clones were selected for sequencing (87058.6, 87058.8, 87058.16). The sequences obtained (SEQ ID NOS:8-10, respectively) were aligned using the DNASIS Multiple sequence alignment program and are shown in Figures 8A through 8F. Clone 87058.6 initiated at base 644 of the published sequence (HUMCATHB, SEQ ID NO:6), clone 87058.8 initiated at base 353 of the published sequence and clone 87058.16 initiated at base 58 of the published sequence, the original clone (87058, SEQ ID NO:7) initiated at base 1058 of the published sequence.

Figures 8A through 8F show an alignment of the obtained sequences with the published human Hsp 90 nucleotide sequence. Clone 87058.16 contains part of the 5'UT and therefore the full coding region of the gene.

Example 3

In Example 3, a full length cDNA (Seq ID NO 11) of a novel P2U purinergic receptor homolog was obtained by the inventive method and is the subject of U.S. Patent Application 08/459,046 filed June 2, 1995, which is hereby incorporated by reference. Inherit™ and BLAST search and alignment tools were used to relate a partial sequence found in Incyte Clone 179696 from the placental cDNA library to the GenBank sequence of RNU09402, a G-protein coupled surface receptor from rat (Rice WR et al (1995) Am J Respir Cell Molec Biol 12:27-32).

The cDNA of Incyte 179696 was extended to full length using a modified XL-PCR (Perkin Elmer) procedure. Primers were designed based on known sequence; one primer was synthesized to initiate extension in the antisense direction (XLR) and the other to extend sequence in the sense direction (XLF). The primers allowed the sequence to be extended "outward" from the known sequence, thus generating amplicons containing new, unknown nucleotide sequence comprising the gene of interest. The primers were designed using Oligo 4.0 (National Biosciences Inc, Plymouth MN) to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

The cDNA library was used as a template, and XLR (bases 278-298) and XLF (bases 587-610) primers were used to extend and amplify the 179696 sequence. By following the instructions for the XL-PCR kit and thoroughly mixing the enzyme, high fidelity amplification is obtained. Beginning with 25 pMol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the MJ PTC200 thermocycler (MJ Research, Watertown MA) and the following parameters:

Step 1	94° C for 60 sec (initial denaturation)
Step 2	94° C for 15 sec
Step 3	65° C for 1 min

Step 4	68° C for 7 min
Step 5	Repeat step 2-4 for 15 additional cycles
Step 6	94° C for 15 sec
Step 7	65° C for 1 min
Step 8	68° C for 7 min + 15 sec/cycle
Step 9	Repeat step 6-8 for 11 additional cycles
Step 10	72° C for 8 min
Step 11	4° C (and holding)

At the end of 28 cycles, 50 µl of the reaction mix was removed; and the remaining reaction mix was run for an additional 10 cycles as outlined below:

Step 1	94° C for 15 sec
Step 2	65° C for 1 min
Step 3	68° C for (10 min + 15 sec)/cycle
Step 4	Repeat step 1-3 for 9 additional cycles
Step 5	72° C for 10 min

A 5-10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Although all extensions potentially contain a full length gene, some of the largest products or bands were selected and cut out of the gel. Further purification involved using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc, Chatsworth CA). After recovery of the DNA, Klenow enzyme was used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitated religation and cloning.

After ethanol precipitation, the products were redissolved in 13 µl of ligation buffer. Then, 1µl T4-DNA ligase (15 units) and 1µl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 µl of appropriate media) were transformed with 3 µl of ligation mixture and cultured in 80 µl of SOC medium (Sambrook J et al, supra). After incubation for one

hour at 37° C, the whole transformation mixture was plated on Luria Broth (LB)-agar (Sambrook J et al, supra) containing carbenicillin at 25 mg/L. The following day, 12 colonies were randomly picked from each plate and cultured in 150 µl of liquid LB/carbenicillin medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 µl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 µl of each sample was transferred into a PCR array.

For PCR amplification, 15 µl of concentrated PCR reaction mix (1.33X) containing 0.75 units of Taq polymerase, a vector primer and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid and sequenced.

Example 4

In this example, the inventive method was used to obtain a novel full length cDNA from the partial sequence found in Incyte clone 08118 which was found to be somewhat homologous to the GenBank sequence of C5a anaphylatoxin receptor, a G-protein coupled surface receptor from dog (Perret J et al (1995) Biochem

J 288:911-17). Based on the partial cDNA sequence, primers (XLR = GAAAGACAGCCACCACCACCACG and XLF = AGAAAGCAAGGCAGTCCATTCAGG) were designed. Essentially the same method outlined in Example 3 above was used to extend the partial sequence of 8118 to obtain the full length sequence (Seq ID NO:12) of a novel C5a-like receptor homolog which is the subject of a U.S. Patent Application 08/462,355 filed June 5, 1995, and whose disclosure is incorporated by reference.

While the present invention has been described with reference to specific enzymes and sequences, particularly PCR enzyme, and formulations containing such, those skilled in the art understand that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, enzyme, process, process step or steps and still carry out the objective, spirit and scope of the invention. All such modifications are intended to be within the scope of the claims appended hereto.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF INVENTION: IMPROVED METHOD FOR OBTAINING FULL LENGTH cDNA SEQUENCES

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
(B) STREET: 3330 Hillview Avenue
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Filed Herewith

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: US 08/487,112
(B) FILING DATE: 7-JUN-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: US 08/462,355
(B) FILING DATE: 5-JUN-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: US 08/459,046
(B) FILING DATE: 2-JUN-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: US 08/566,334
(B) FILING DATE: 1-DEC-1995

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION SERIAL NO: US 60/006,809
(B) FILING DATE: 15-NOV-1995

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Luther, Barbara J.
(B) REGISTRATION NUMBER: 33954
(C) REFERENCE/DOCKET NUMBER: RP-001-1 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415-855-0555

(B) TELEFAX: 415-852-0195

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2543 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank HUMHSP90
(B) CLONE: Accession No. M16660

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTCCGGCGCA GTGTTGGGAC TGTCTGGGTA TCGGAAAGCA AGCCTACGTT GCTCACTATT 60

ACGTATAATC CTTTTCTTTT CAAGATGCCT GAGGAAGTGC ACCATGGAGA GGAGGAGGTG 120

GAGACTTTTG CCTTTCAGGC AGAAATTGCC CAACTCATGT CCCTCATCAT CAATACCTTC 180

TATTCCAACA AGGAGATTTT CCTTCGGGAG TTGATCTCTA ATGCTTCTGA TGCCTTGGAC 240

AAGATTCGCT ATGAGAGCCT GACAGACCCT TCGAAGTTGG ACAGTGGTAA AGAGCTGAAA 300

ATTGACATCA TCCCCAACCC TCAGGAACGT ACCCTGACTT TGGTAGACAC AGGCATTGGC 360

ATGACCAAAG CTGATCTCAT AAATAATTTG GGAACCATTG CCAAGTCTGG TACTAAAGCA 420

TTCATGGAGG CTCTTCAGGC TGGTGCAGAC ATCTCCATGA TTGGGCAGTT TGGTGTTGGC 480

TTTTATTCTG CCTACTTGGT GGCAGAGAAA GTGGTTGTGA TCAGAAAGCA CAACGATGAT 540

GAACAGTATG CTTGGGAGTC TTCTGCTGGA GGTTCCTTCA CTGTGCGTGC TGACCATGGT 600

GAGCCCATTG GCATGGGTAC CAAAGTGATC CTCCATCTTA AAGAAGATCA GACAGAGTAC 660

CTAGAAGAGA GGCGGGTCAA AGAAGTAGTG AAGAAGCATT CTCAGTTCAT AGGCTATCCC 720

ATCACCCTTT ATTTGGAGAA GGAACGAGAG AAGGAAATTA GTGATGATGA GGCAGAGGAA 780

GAGAAAGGTG AGAAAGAAGA GGAAGATAAA GATGATGAAG AAAAGCCCAA GATCGAAGAT 840

GTGGGTTCAG ATGAGGAGGA TGACAGCGGT AAGGATAAGA AGAAGAAAAC TAAGAAGATC 900

AAAGAGAAAT ACATTGATCA GGAAGAACTA AACAAGACCA AGCCTATTTG GACCAGAAAC 960

CCTGATGACA TCACCCAAGA GGAGTATGGA GAATTCTACA AGAGCCTCAC TAATGACTGG 1020

GAAGACCACT TGGCAGTCAA GCACTTTTCT GTAGAAGGTC AGTTGGAATT CAGGGCATTG 1080

CTATTTATTC CTCGTCGGGC TCCCTTTGAC CTTTTTGAGA ACAAGAAGAA AAAGAACAAC 1140

ATCAAACTCT ATGTCCGCCG TGTGTTCATC ATGGACAGCT GTGATGAGTT GATACCAGAG 1200

TATCTCAATT TTATCCGTGG TGTGGTTGAC TCTGAGGATC TGCCCCTGAA CATCTCCCGA 1260

GAAATGCTCC AGCAGAGCAA AATCTTGAAA GTCATTCGCA AAAACATTGT TAAGAAGTGC 1320

CTTGAGCTCT TCTCTGAGCT GGCAGAAGAC AAGGAGAATT ACAAGAAATT CTATGAGGCA 1380

TTCTCTAAAA ATCTCAAGCT TGGAATCCAC GAAGACTCCA CTAACCGCCG CCGCCTGTCT 1440

GAGCTGCTGC GCTATCATAC CTCCCAGTCT GGAGATGAGA TGACATCTCT GTCAGAGTAT 1500

GTTTCTCGCA TGAAGGAGAC ACAGAAGTCC ATCTATTACA TCACTGGTGA GAGCAAAGAG 1560

CAGGTGGCCA ACTCAGCTTT TGTGGAGCGA GTGCGGAAAC GGGGCTTCGA GGTGGTATAT 1620

ATGACCSAGC CCATTGACGA GTACTGTGTG CAGCAGCTCA AGGAATTTGA TGGGAAGAGC 1680

CTGGTCTCAG TTACCAAGGA GGGTCTGGAG CTGCCTGAGG ATGAGGAGGA GAAGAAGAAG 1740

ATGGAAGAGA GCAAGGCAAA GTTTGAGAAC CTCTGCAAGC TCATGAAAGA AATCTTAGAT 1800

AAGAAGGTTG AGAAGGTGAC AATCTCCAAT AGACTTVTGT CTTCACCTTG CTGCATTGTG 1860

ACCAGCACCT ACGGCTGGAC AGCCAATATG GAGCGGATCA TGAAAGCCCA GGCACTTCGG 1920

GACAACTCCA CCATGGGCTA TATGATGGCC AAAAAGCACC TGGAGATCAA CCCTGACCAC 1980

CCCATTGTGG AGACGCTGCG GCAGAAGGCT GAGGCCGACA AGAATGATAA GGCAGTTAAG 2040

GACCTGGTGG TGCTGCTGTT TGAAACCGCC CTGCTATCTT CTGGCTTTTC CCTTGAGGAT 2100

CCCCAGACCC ACTCCAACCG CATCTATGGC ATGATCAAGC TAGGTCTAGG TATTGATGAA 2160

GATGAAGTGG CAGCAGAGGA ACCCAATGCT GCAGTTCCTG ATGAGATCCC CCCTCTCGAG 2220

GGCGATGAGG ATGCGTCTCG CATGGAAGAA GTCGATTAGG TTAGGAGTTC ATAGTTGGAA 2280

AACTTGTGCC CTTGTATAGT GTCCCCATGG GCTCCCACTG CAGCTTCGAG TGCCCCTGTC 2340

CCACCTGGCT CCCCCTGCTG GTGTCTAGTG TTTTTTTCCC TCTCCTGTCC TTGTGTTGAA 2400

GGCAGTAAAC TAAGGGTGTC AAGCCCCATT CCCTCTCTAC TCTTGACAGC AGGATTGGAT 2460

GTTGTGTATT GTGGTTTATT TTATTTTCTT CATTTTGTTC TGAAATTAAA GTATGCAAAA 2520

TAAAGAATAT GCCGTTTTTA TAC 2543

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 261 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: THP-1
(B) CLONE: 14201

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

AAGAAAAAGA ACAACATCAA ACTCTATGTC CGCCGTGTGT TCATCATGGC AGCTGTGATG 60

AGTTGATACC AGAGTATCTC AATTTTATCC GTGGTGTGGT TGACTTGAGG TCTGCCCCTG 120

AACATCTCCC GGAAATGCTC CAGCAGAGCA AAATCTTGAA AGGCATTCGC AAAAACATTG 180

TTAAGAGTGC CTTAGCTCTT CTCTAGCTGG CAGAAGCAAG GGGATTTCAA GAAATTCTTT 240

TGGGGGGATT TCTTAAAAAT T 261

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: THP-1
(B) CLONE: 14201.3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCTGGGTATC GGAAAGCAAG CCTACGTTGC TCACTATTAC GTATAATCCT TTTCTTCAAG 60

ATGCCTGAGG AAGTGCACCA TGGAGAGGAG GAGGTGGAGA CTTTTGCCTT TCAGGCAGAA 120

ATTGCCCAAC TCATGTCCCT CATCATCAAT ACCTCCTATT CCAACAAGGA GATTTCCTCG 180

GGAGTTGATC TCTAATGCTT CTGATGCCTC GGACAAGATT CGCTATGAAG CCTGACAGAC 240

CCTTCGAAGT GGTCAGCGGC AAGAGCTGAA AATTGACATC ATCCCCAACC CTCAGGAACG 300

TCCCTGTACT TTGGGTAGAC ACAGGCATTG GCATAAACAA AGCTGACCTC ATATTATTCG 360

GGGAACCATT GCCAAGTCTT GTCTAAAAGC ATTCATGGAG GCTCTCAGGT TGGCGCAGAC 420

ATCTCCAGAT TGGCAGGTGG GTGTTGGCTT TATTCTGCCC ACTTGGTGGC AGAGAAAT 478

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: THP-1
(B) CLONE: 14201.5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTTGGGACTG TCTGGGTATC GGAAAGCAAG CCTACGTTGC TCACTATTAC GTATAATCCT 60

TTTCTTTTCA AGATGCCTGA GGAAGTGCAC CATGGAGAGG AGGAGGTGGA GACTTTTGCC 120

TTTCAGGCAG AAATTGCCCA ACTCATGTCC CTCATCATCA ATACCTCCTA TTCCAACAAG 180

GAGATTTTCC TTCGGGAGTT GATCTCTAAT GCTTCTGATG CCTTGGACAA GATTCGCTAT 240

GAGAGCCTGA CAGACCCTTC GAAGTTGGAC AGTGGTAAAG AGCTGAAAAT TGACATCATC 300

CCCAACCCTC AGGAACGTAC CCTGACTTTG GGTAGACACA GGCATCGGCA TGACCAAAAG 360

CTGATCTCAT AATAATTGGG AACCATTGCA AGTCTGGTAC TAAAGCATTC ATGGAGGCTC 420

TTCAGGCTGG TGCAGACATC TCCATGATTG GGCAGCTTGG GTGTTGCTTT ATTCTGCCTC 480

CTTGGTGGCA GAGAAAGTGT TGTGATCA 508

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 547 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: THP-1
(B) CLONE: 14201.13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTGAGAGTAT GTCGAGTTAC TGTGGAGGTT CCTTCACTGC GTGCTGACAT GGTGAGCCCA 60

TGGGAGCGGT ACCAAGTGAT CCTCCATCTC AAAGAAGATC AGACAGAGTA CCTAGAGAGA 120

GGCGGATCAA AGAGTAGTGA TGAGCATCCT CAGATCATAG GCTATCCCAT CACCCTTTTT 180

TGGAGAAGGA CGAGAGAAGG AATTAGGATG ATGAGGCAGA GGAAGAGAAT GGTGAGAATG 240

AAGAGGAGTA ACGATGATGA AGAAACCCCA AGATCGATGA TGTGGTTCAG ATGAGGGAT 300

GACAGCGGTA GATAAGAAGA AGAAACTAGA ATCATCGGAT CATGACAGGA AGAACTAACA 360

GATCATCTTT CGGCCAGAAT CCCTGATGTC ATCACCCAAG AGGGTATGGA GATTTCTACA 420

TGCAGCTCAC TTTACTGGGC AAGACACTTG GCAGCAACAC TTTTCTGTAG AAGGCCATTG 480

CATCACGCAT TGCTATTCTT CCCTGGCCGT CTCCTTTGAC CTGGTCTGGC ATCATGGTGT 540

CTTGATC 547

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1996 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank HUMCATHB
(B) CLONE: Accession No. L16510

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TCCGGCAACG CCAACCGCTC CGCTGCGCGC AGGCTGGGCT GCAGGCTCTC GGCTGCAGCG 60

CTGGGCTGGT GTGCAGTGGT GCGACCACGG CTCACGGCAG CCTCAGCCAC CCAGATGTAA 120

GCGATCTGGT TCCCACCTCA GCCTCCCGAG TAGTGGATCT AGGATCCGGC TTCCAACATG 180

TGGCAGCTCT GGGCCTCCCT CTGCTGCCTG CTGGTGTTGG CCAATGCCCG GAGCAGGCCC 240

TCTTTCCATC CCCTGTCGGA TGAGCTGGTC AACTATGTCA ACAAACGGAA TACCACGTGG 300

CAGGCCGGGC ACAACTTCTA CAACGTGGAC ATGAGCTACT TGAAGAGGCT ATGTGGTACC 360

TTCCTGGGTG GGCCCAAGCC ACCCCAGAGA GTTATGTTTA CCGAGGACCT GAAGCTGCCT 420

GCAAGCTTCG ATGCACGGGA ACAATGGCCA CAGTGTCCCA CCATCAAAGA GATCAGAGAC 480

CAGGGCTCCT GTGGCTCCTG CTGGGCCTTC GGGGCTGTGG AAGCCATCTC TGACCGGATC 540

TGCATCCACA CCAATGCGCA CGTCAGCGTG GAGGTGTCGG CGGAGGACCT GCTCACATGC 600

TGTGGCAGCA TGTGTGGGGA CGGCTGTAAT GGTGGCTATC CTGCTGAAGC TTGGAACTTC 660

TGGACAAGAA AAGGCCTGGT TTCTGGTGGC CTCTATGAAT CCCATGTAGG GTGCAGACCG 720

TACTCCATCC CTCCCTGTGA GCACCACGTC AACGGCTCCC GGCCCCCATG CACGGGGGAG 780

GGAGATACCC CCAAGTGTAG CAAGATCTGT GAGCCTGGCT ACAGCCCGAC CTACAAACAG 840

GACAAGCACT ACGGATACAA TTCCTACAGC GTCTCCAATA GCGAGAAGGA CATCATGGCC 900

GAGATCTACA AAAACGGCCC CGTGGAGGGA GCTTTCTCTG TGTATTCGGA CTTCCTGCTC 960

TACAAGTCAG GAGTGTACCA ACACGTCACC GGAGAGATGA TGGGTGGCCA TGCCATCCGC 1020

ATCCTGGGCT GGGGAGTGGA GAATGGCACA CCCTACTGGC TGGTTGCCAA CTCCTGGAAC 1080

ACTGACTGGG GTGACAATGG CTTCTTTAAA ATACTCAGAG GACAGGATCA CTGTGGAATC 1140

GAATCAGAAG TGGTGGCTGG AATTCCACGC ACCGATCAGT ACTGGGAAAA GATCTAATCT 1200

GCCGTGGGCC TGTCGTGCCA GTCCTGGGGG CGAGATCGGG GTAGAAATGC ATTTTATTCT 1260

TTAAGTTCAC GTAAGATACA AGTTTCAGGC AGGGTCTGAA GGACTGGATT GGCCAAACAT 1320

CAGACCTGTC TTCCAAGGAG ACCAAGTCCT GGCTACATCC CAGCCTGTGG TTACAGTGCA 1380

GACAGGCCAT GTGAGCCACC GCTGCCAGCA CAGAGCGTCC TTCCCCCTGT AGACTAGTGC 1440

CGTGGGAGTA CCTGCTGCCC AGCTGCTGTG GCCCCCTCCG TGATCCATCC ATCTCCAGGG 1500

AGCAAGACAG AGAGGCAGGA TGGAAAGCGG AGTTCCTAAC AGGATGAAAG TTCCCCCATC 1560

AGTTCCCCCA GTACCTCCAA GCAAGTAGCT TTCCACATTT GTCACAGAAA TCAGAGGAGA 1620

GATGGTGTTG GGAGCCCTTT GGAGAACGCC AGTCTCCAGG TCCCCCTGCA TCTATCGAGT 1680

TTGCAATGTC ACAACCTCTC TGATCTTGTG CTCAGCATGA TTCTTTAATA GAAGTTTTAT 1740

TTTTCGTGCA CTCTGCTAAT CATGTGGGTG AGCCAGTGGA ACAGCGGGAG CCTGTGCTGG 1800

TTTGCAGATT GCCTCCTAAT GACGCGGCTC AAAAGGAAAC CAAGTGGTCA GGAGTTGTTT 1860

CTGACCCACT GATCTCTACT ACCACAAGGA AAATAGTTTA GGAGAAACCA GCTTTTACTG 1920

TTTTTGAAAA ATTACAGCTT CACCCTGTCA AGTTAACAAG GAATGCCTGT GCCAATAAAA 1980

GGTTTCTCCA ACTTGA 1996

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: LIVER
(B) CLONE: 87058

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGGCACGAGC CAACTCCTGG AACACTGACT GGGGTGACAA TGGCTTCTTT AAAATACTCA 60

GAGGACAGGT TCACTGTGGA ATCGAATCAG AAGTGGTGGC TGGAATTCCA CGCACCGTTC 120

AGTACTGGGA AAAGTCTAAT CTGCCGTGGG CCTTCGTGCC AGTCCTGGGG GCGAGATGGG 180

GGTAGAAATG CATTTTATTC TTTAAGTTCA CGTAAGATAC AAGTTTCAGA CAGGGGTCTA 240

AGGCCTGGTT GCCAAAATCA GACCTGTTTT TCAAGGGGCC CAAGTCCTGG GTTC 294

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 552 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Liver
(B) CLONE: 87058.6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTGAAGCTTG GAACTTCTGG ACAAGAAAAG GCCTGGTTTC TGGTGGCCTC TATGAATCCC 60

ATGTAGGGTG CAGACCGTAC TCCATCCCTC CCTGTGAGCA CCACGTCAAC GGCTCCCGGC 120

CCCCATGCAC GGGGGAGGGA GATACCCCCA AGTGTAGCAA GATCTGTGAG CCTGGCTACA 180

GCCCGACCTA CAAACAGGAC AAGCACTACG GATACAATTC CTACAGCGTC TCCAATAGCG 240

AGAAGGACAT CATGGCCGAG ATCTACAAAA ACGGCCCCGT GGAGGGAGCT TTCTCTGTGT 300

ATTCGGACTT CCTGCTCTAC AAGTCAGGAG TGTACCAACA CGTCACCGGA GAGATGATGG 360

GTGGCCATGC CATCCGCATC CTGGGCTGGG GAGTGGAGAA TGGCACAACC TACTGGCTGG 420

TTGGCAACTC CTGGAACACT GACTGGGGTG ACAATGGGTT CACTGTGGAA TCGAATCAGA 480

AGTGGTGGTG GAATTCCACG CACGATCAAG TGCTGGGAAA AGATCTTAAT CTGCCGGGGC 540

TGTCGGCCAG TC 552

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 559 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Liver
(B) CLONE: 87058.8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAGGTACCTT CCTGGGTGGG CCCAAGCCAC CCCAGAGAGT TATGTTTACC GAGGACCTGA 60

AGCTGCCTGC AAGCTTCGAT GCACGGGAAC AATGGCCACA GTGTCCCACC ATCAAAGAGA 120

TCAGAGACCA GGGTCCTGTG GCTCCTGCTG GGCCTTCGGG GCTGTGGAAG CCATCTCTGA 180

CCGGATCTGA TCCACACCAA TGCGCACGTC AGCGTGGAGG TGTCGGCGGA GGACTGCTCA 240

CATGCTGTGG CAGATGTGTG GGGACGGCTG TAATGGTGGC TATCCTGCTG AAGCTTGGAC 300

TTCTGGACAA GAAAAGGCCC TGGTTTCTGG TGGCCTCTAT GATCCCATGT AGGGTGTAGA 360

CCGTACTCCA TCCCTCCCTG TGAAGCACCA CGTCAACGGT TCCCGGGCCC CATGCACGGG 420

GAGGGAGATA CCCCCAAGTG TAACAAGATC TGTGAGCCTG GGTACAGTCC CGACCACAAA 480

CAGGAAAAGC ACTACGGATA CAATTCCTCA GGTCTCCAAT AGTGAGAAGG GACATCATGC 540

CGAGATCTAC AATAACGGC 559

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Liver
(B) CLONE: 87058.16

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CGGTTGAGAT TCGGACAGTC CGAAAACGTC CGGCAAGTCA CCCGCTCCGC TGGCGCAGGC 60

TGGGTGCAGG CTCTCGGTGC AGGCTGGGTG GATCTAGGAT CCGGCTTCCA ACATGTGGCA 120

GTTCTGGGCC TCCCTCTGTG CCTGCTGGTG TTGGACAATG CCCGGAGGAG GCCTCTTTCC 180

ATCCCCTGTC GGATGAGCTG GTCACTATGT CAACAAACGG AATACCACGT GGAGGCCGGG 240

AACAACTTCT ACAACGTGGA CATGAGCTAC TTGAGAGGTA TGTGGTACCT TCCTGGGTGG 300

GCCCAAGCCA CCCCAGAGAG TTTGTTTACC GAGGACCTGA GCTGCCTGCA AGCTTCGAAG 360

GACGGGAACA ATGGCCACAG TGTCCCACCA TCAAAGAGAT CAGAGACAGG GCTCCTGTGG 420

TCCTGCTGGG CCTCCGGGGC TGTGGAAGCA TCTCTGACCG GATCTGCATC CACACCAATG 480

GCACGTCAGC GTGGTGGTGT CGGGGAAGAC CTGATCACCT TTGTGGTAGC ATGTGTGGGG 540

GACGGCTGTA ATGGTGGTTA TCCTGTGAAG CTGGGCCTTC TAGAAAGAAA AGGCTGTTTT 600

GGTGGCCTTA TGACTCCCAT GT 622

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 984 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: Placenta
(B) CLONE: 179696

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATGGAATGGG ACAATGGCAC AGACCAGGCT CTGGGCTTGC CACCCACCAC CTGTGTCTAC 60

CGCGAGAACT TCAAGCAACT GCTGCTCCCA CCTGTGTATT CGGCGGTGCT GGCGCCTGCC 120

CTCCCGCTGA ACATCTGTGT CATTACCCAG ATCTGCACGT CCCGCCGGGC CCTGACCCGC 180

ACGGCCGTGT ACACCCTAAA CCTTGCTCTG GCTGACCTGC TATATGCCTG CTCCCTGCCC 240

CTGCTCATCT ACAACTATGC CCAAGGTGAT CACTGGCCCT TTGGCGACTT CGCCTGCCGC 300

CTGGTCCGCT TCCTCTTCTA TGCCAACCTG CACGGGAGGA TCCTCTTCCT CACCTGCATC 360

AGCTTCCAGC GCTACCTGGG CATCTGCCAC CCGCTGGCCC CCTGGCACAA ACGTGGGGGC 420

CGCCGGGCTG CCTGGCTAGT GTGTGTAGCC GTGTGGCTGG CCGTGACAAC CCAGTGCCTG 480

CCCACAGCCA TCTTCGCTGC CACAGGCATC CAGCGTAACC GCACTGTCTG TTATGACCTC 540

AGCCCGCCTG CCCTGGCCAC CCACTATATG CCCTATGGGA TGGCTCTCAC TGTCATCGGC 600

TTCCTGCTGC CCTTTGCTGC CCTGCTGGCC TGCTACTGTC TCCTGGCCTG CCGCCTGTGC 660

CGCCAGGATG GCCCGGCAGA GCCTGTGGCC CAGGAGCGGC GTGGCAAGGC GGCCCGCATG 720

GCCGTGGTGG TGGCTGCTGT CTTTGGCATC AGCTTCCTGC CTTTTCACAT CACCAAGACA 780

GCCTACCTGG CAGTGCGCTC GACGCCGGGC GTCCCCTGCA CTGTATTGGA GGCCTTTGCA 840

GCGGCCTACA AAGGCACGCG GCCGTTTGCC AGTGCCAACA GCGTGCTGGA CCCCATCCTC 900

TTCTACTTCA CCCAGAAGAA GTTCCGCCGG CGACCACATG AGCTCCTACA GAAACTCACA 960

GACAAATGGC AGAGGCAGGG TCGC 984

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1446 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:

(A) LIBRARY: Mast Cell
(B) CLONE: 6118

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

ATGGCGTCTT TCTCTGCTGA GACCAATTCA ACTGACCTAC TCTCACAGCC ATGGAATGAG 60

CCCCCAGTAA TTCTCTCCAT GGTCATTCTC AGCCTTACTT TTTTACTGGG ATTGCCAGGC 120

AATGGGCTGG TGCTGTGGGT GGCTGGCCTG AAGATGCAGC GGACAGTGAA CACAATTTGG 180

TTCCTCCACC TCACCTTGGC GGACCTCCTC TGCTGCCTCT CCTTGGCCTT CTCGCTGGCT 240

CACTTGGCTC TCCAGGGACA GTGGCCCTAC GGCAGGTTCC TATGCAAGCT CATCCCCTCC 300

ATCATTGTCC TCAACATGTT TGGCAGTGTC TTCCTGCTTA CTGCCATTAG CCTGGATCGC 360

TGTCTTGTGG TATTCAAGCC AATCTGGTGT CAGAATCATC GCAATGTAGG GATGGCCTGC 420

TCTATCTGTG GATGTATCTG GGTGGTGGCT TTTGTGTTGT GCATTCCTGT GTTCGTGTAC 480

CGGGAAATCT TCACTACAGA CAACCATAAT AGATGTGGCT ACAAATTTGG TCTCTCCAGC 540

TCATTAGATT ATCCAGACTT TTATGGGGAT CCACTAGAAA ACAAGTCTCT TGAAAACATT 600

GTTCAGCCGC CTGGAGAAAT GAATGATAGG TTAGATCCTT CCTCTTTCCA AACAAATGAT 660

CATCCTTGGA CAGTCCCCAC TGTCTTCCAA CCTCAAACAT TTCAAAGACC TTCTGCAGAT 720

TCACTCCCTA GGGGTTCTGC TAGGTTAACA AGTCAAAATC TGTATTCTAA TGTATTTAAA 780

CCTGCTGATG TGGTCTCACC TAAAATCCCC AGTGGGTTTC CTATTGAAGA TCACGAAACC 840

AGCCCACTGG ATAACTCTGA TGCTTTTCTC TCTACTCATT TAAAGCTGTT CCCTAGCGCT 900

TCTAGCAATT CCTTCTACGA GTCTGAGCTA CCACAAGGTT TCCAGGATTA TTACAATTTA 960

GGCCAATTCA CAGATGACGA TCAAGTGCCA ACACCCCTCG TGGCAATAAC GATCACTAGG 1020

CTAGTGGTGG GTTTCCTGCT GCCCTCTGTT ATCATGATAG CCTGTTACAG CTTCATTGTC 1080

TTCCGAATGC AAAGGGGCCG CTTCGCCAAG TCTCAGAGCA AAACCTTTCG AGTGGCCGTG 1140

GTGGTGGTGG CTGTCTTTCT TGTCTGCTGG ACTCCATACC ACATTTGGGG AGTCCTGTCA 1200

TTGCTTACTG ACCCAGAAAC TCCCTTGGGG AAAACTCTGA TGTCCTGGGA TCATGTATGC 1260

ATTGCTCTAG CATCTGCCAA TAGTTGCTTT AATCCCTTCC TTTATGCCCT CTTGGGGAAA 1320

GATTTTAGGA AGAAAGCAAG GCAGTCCATT CAGGGAATTC TGGAGGCAGC CTTCAGTGAG 1380

GAGCTCACAC GTTCCACCCA CTGTCCCTCA AACAATGTCA TTTCAGAAAG AAATAGTACA 1440

ACTGTG 1446

CLAIMS

1. A method of extending the sequence of a partial complementary DNA (cDNA) using polymerase chain reaction (PCR), comprising the steps of:

a) combining a first and second PCR primer with nucleic acid from a cDNA library expected to contain said partial cDNA, or a genomic library, under conditions suitable for synthesis of nucleic acid PCR products from the first and second primers, wherein said first and second primers are capable of annealing to opposite strands of the partial cDNA or genomic DNA and initiating nucleic acid synthesis in an outward manner and wherein the first primer is capable of being extended by DNA polymerase in an antisense direction and the second primer is capable of being extended in a sense direction.

b) purifying the PCR products, and

c) identifying extended nucleotide sequences derived from said partial cDNA or said genomic DNA.

2. The method of Claim 1 wherein identifying extended sequences comprises nucleic acid sequencing.

3. The method of Claim 2 further comprising extending the nucleotide sequences of step 6c by repeating steps 6a through 6c on the nucleotide sequences identified in step 6c.

4. A method of extending the nucleotide sequence of a partial complementary DNA (cDNA) using polymerase chain reaction (PCR), comprising the steps of:

a) combining a first and second PCR primer with nucleic acid from a cDNA library expected to contain said partial cDNA, or a genomic library, under conditions suitable for synthesis of nucleic acid PCR products from the first and second primers, wherein said first and second primers are capable of annealing to opposite strands of the partial cDNA or genomic DNA and initiating nucleic acid synthesis in an outward manner and wherein the first primer is capable of being extended by DNA polymerase in an

antisense direction and the second primer is capable of being extended in a sense direction.

b) purifying the PCR products,

c) ligating the purified PCR products under conditions suitable for the formation of circular closed nucleic acid,

d) transforming a host cell with the circular closed nucleic acid and culturing the transformed host cell under conditions suitable for growth,

e) recovering said circular closed nucleic acid from the cultured, transformed host cell,

f) identifying extended nucleotide sequences derived from said partial cDNA or said genomic DNA.

5. The method of Claim 4 wherein identifying extended sequences compriass nucleic acid sequencing.

6. The method of Claim 4 wherein culturing the transformed host cell under conditions suitable for growth comrpises culturing in the presence of selective antibiotic conditions.

7. The method of Claim 4 wherein said host cell is E.coli.

8. The method of Claim 4 wherein after step 4b and prior to step 4c, the purified PCR products are treated under conditions sutiable for converting nucleic acid overhangs to blunt ends.

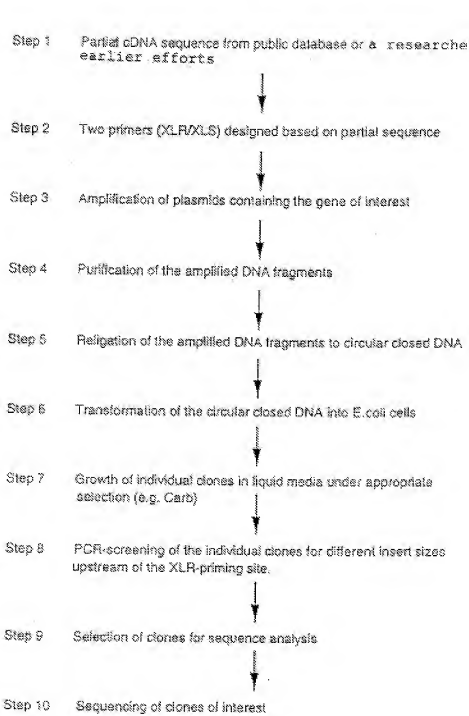

Step 1
Partial cDNA sequence from public database or a researcher's earlier efforts
Step 2
Two primers (XLR/XLS) designed based on partial sequence
Step 3
Amplification of plasmids containing the gene of interest
Step 4
Purification of the amplified DNA fragments
Step 5
Religation of the amplified DNA fragments to circular closed DNA
Step 6
Transformation of the circular closed DNA into E.coli cells
Step 7
Growth of individual clones in liquid media under appropriate selection (e.g. Carb)
Step 8
PCR-screening of the individual clones for different insert sizes upstream of the XLR-priming site.
Step 9
Selection of clones for sequence analysis
Step 10
Sequencing of clones of interest

FIGURE 1

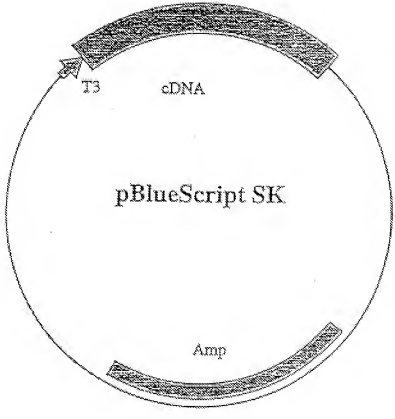


FIGURE 2

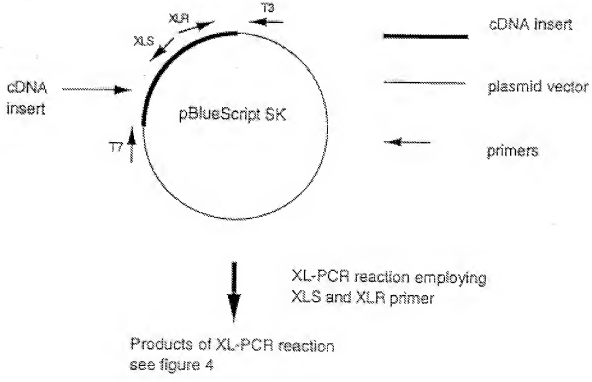

T3
XLR
XLS
cDNA insert
cDNA insert
pBlueScript SK
T7
plasmid vector
primers
XL-PCR reaction employing
XLS and XLR primer
Products of XL-PCR reaction
see figure 4

FIGURE 3

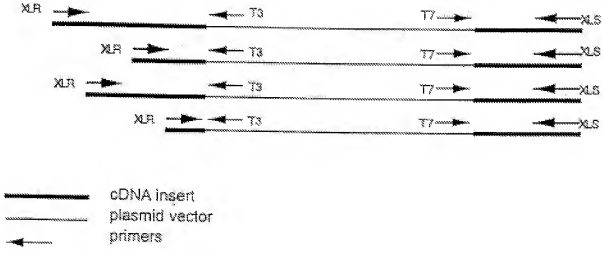


FIGURE 4

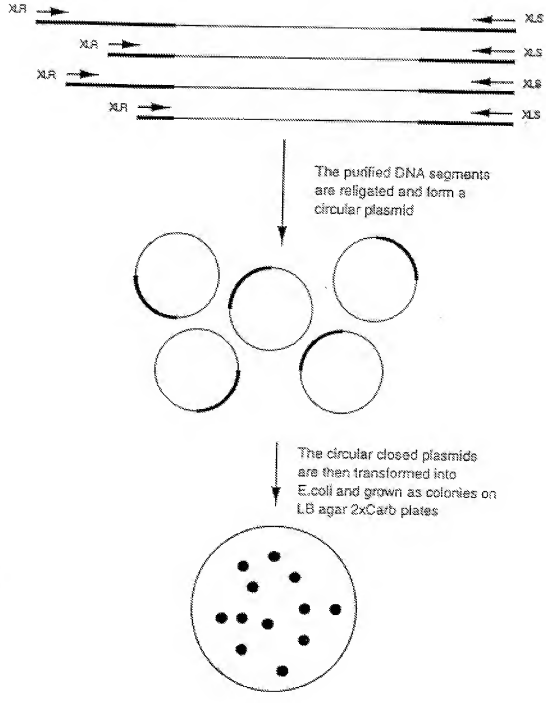

XLR
XLS
XLR
XLS
XLR
XLS
XLR
XLS
The purified DNA segments are religated and form a circular plasmid
The circular closed plasmids are then transformed into E.coli and grown as colonies on LB agar 2xCarb plates

FIGURE 5

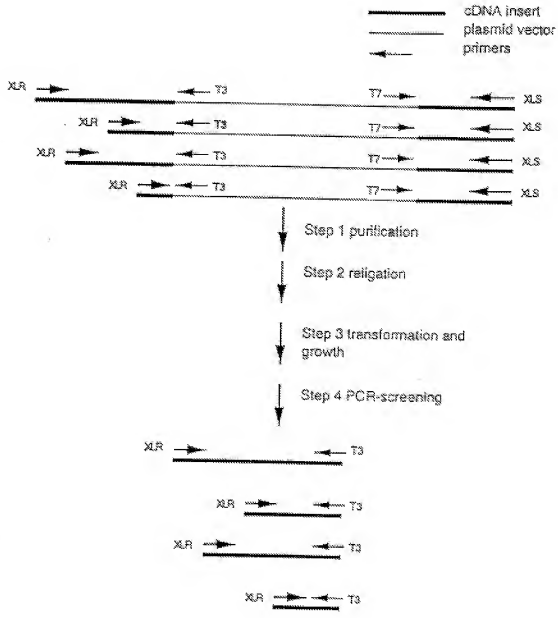

cDNA insert
plasmid vector
primers
XLR
T3
T7
XLS
Step 1 purification
Step 2 religation
Step 3 transformation and growth
Step 4 PCR-screening

FIGURE 6

```
                      10         20         30         40         50
Hsp 90           1 CTCCGGCGCA GTGTTGGGAC TGTCTGGGTA TCGGAAAGCA AGCCTACGTT   50
14201            1 ---------- ---------- ---------- ---------- ----------   50
14201.3          1 ---------- ---------- --GCTGGGTA TCGGAAAGCA AGCCTACGTT   50
14201.5          1 ---------- --GTTGGGAC TGTCTGGGTA TCGGAAAGCA AGCCTACGTT   50
14201.13         1 ---------- ---------- ---------- ---------- ----------   50

                      60         70         80         90        100
Hsp 90          51 GCTCACTATT ACGTATAATC CTTTTCTTTT CAAGATGCCT GAGGAAGTGC  100
14201           51 ---------- ---------- ---------- ---------- ----------  100
14201.3         51 GCTCACTATT ACGTATAATC CTTTTCTNTN CAAGATGCCT GAGGAAGTGC  100
14201.5         51 GCTCACTATT ACGTATAATC CTTTTCTTTT CAAGATGCCT GAGGAAGTGC  100
14201.13        51 ---------- ---------- ---------- ---------- ----------  100

                     110        120        130        140        150
Hsp 90         101 ACCATGGAGA GGAGGAGGTG GAGACTTTTG CCTTTCAGGC AGAAATTGCC  150
14201          101 ---------- ---------- ---------- ---------- ----------  150
14201.3        101 ACCATGGAGA GGAGGAGGTG GAGACTTTTG CCTTTCAGGC AGAAATTGCC  150
14201.5        101 ACCATGGAGA GGAGGAGGTG GAGACTTTTG CCTTTCAGGC AGAAATTGCC  150
14201.13       101 ---------- ---------- ---------- ---------- ----------  150

                     160        170        180        190        200
Hsp 90         151 CRACTCATGT CCCTCATCAT CAATACCTTC TATTCCAACA AGGAGATTTT  200
14201          151 ---------- ---------- ---------- ---------- ----------  200
14201.3        151 CRACTCATGT CCCTCATCAT CAATACCTCC TATTCCAACA AGGAGATTNT  200
14201.5        151 CRACTCATGT CCCTCATCAT CAATACCTCC TATTCCAACA AGGAGATTTT  200
14201.13       151 ---------- ---------- ---------- ---------- ----------  200

                     210        220        230        240        250
Hsp 90         201 CCTTCGGGAG TTGATCTCTA ATGCTTCTGA TGCCTTGGAC AAGATTCGCT  250
14201          201 ---------- ---------- ---------- ---------- ----------  250
14201.3        201 CCTNCGGGAG TTGATCTCTA ATGCTTCTGA TGCCTGGGAC AAGATTCGCT  250
14201.5        201 CCTTCGGGAG TTGATCTCTA ATGCTTCTGA TGCCTTGGAC AAGATTCGCT  250
14201.13       201 ---------- ---------- ---------- ---------- ----------  250

                     260        270        280        290        300
Hsp 90         251 ATGAGAGCCT GACAGACCCT TCGAAGTTGG ACAGTGGTAA AGAGCTGAAA  300
14201          251 ---------- ---------- ---------- ---------- ----------  300
14201.3        251 ATGANAGCCT GACAGACCCT TCGAAGTNGG TCAGCGGCAA NGAGCTGAAA  300
14201.5        251 ATGAGAGCCT GACAGACCCT TCGAAGTTGG ACAGTGGTAA AGAGCTGAAA  300
14201.13       251 ---------- ---------- ---------- ---------- ----------  300
```

FIGURE 7A

```
                      310         320         330         340         350
Hsp 90          301 ATTGACATCA TCCCCAACCC TCAGGAACGT ACCCTGACTT TGGTAGACAC  350
14201           301 ---------- ---------- ---------- ---------- ----------  350
14201.3         301 ATTGACATCA TCCCCAACCC TCAGGAACGT NCCCTGACTT TGGTAGACAC  350
14201.5         301 ATTGACATCA TCCCCAACCC TCAGGAACGT ACCCTGACTT TGGTAGACAC  350
14201.13        301 ---------- ---------- ---------- ---------- ----------  350

                      360         370         380         390         400
Hsp 90          351 AGGCATTGGC ATGACCAAAG CTGATCTCAT AAaTAATTtG GGAACCATTG  400
14201           351 ---------- ---------- ---------- ---------- ----------  400
14201.3         351 AGGCATTGGC ATGAaaCAAG CTGAcCTCAT WAnTTAATTnG GGgAaCCaTt  400
14201.5         351 AGGCATcGGC ATGACCAAAG CTGATCTCAT AAnTAATTnG GGAACCATTG  400
14201.13        351 ---------- ---------- ---------- ---------- ----------  400

                      410         420         430         440         450
Hsp 90          401 CCAAGTCTGG TACTAAAGCA TTCATGGAGG CTCTTCAGGC TGGTGCAGAC  450
14201           401 ---------- ---------- ---------- ---------- ----------  450
14201.3         401 CCAAGTCTTG TNCTAAAGCA TTCATGGAGG CTCTNCAGGN TGGcGCAGAC  450
14201.5         401 NCAAGTCTGG TACTAAAGCA TTCATGGAGG CTCTTCAGGC TGGTGCAGAC  450
14201.13        401 ---------- ---------- ---------- ---------- ----------  450

                      460         470         480         490         500
Hsp 90          451 ATCTCCATGA TTGGGCAGTT LGGTGTTGGC TttTATTCTG CCTACTTGGT  500
14201           451 ---------- ---------- ---------- ---------- ----------  500
14201.3         451 ATCTCCANGA TTNGGCAGNT GGGTGTTGGC TTNTATTCTG CCnACTTGGT  500
14201.5         451 ATCTCCATGA TTGGGCAGTT GGGTGTTGNC TTnTATTCTG CCTnCTTGGT  500
14201.13        451 ---------- ---------- ---------- ---------- ----------  500

                      510         520         530         540         550
Hsp 90          501 GGCAGAGAAA GTGGTTGTGA TCAGAAAGCA CAACGATGAT GAacAGTATG  550
14201           501 ---------- ---------- ---------- ---------- ----------  550
14201.3         501 GGCAGAGAAA NNT....... .......... .......... ..........  550
14201.5         501 GGCAGAGAAA GTNGTTGTGA TCA....... .......... .......... ..........  550
14201.13        501 ---------- ---------- ---------- --------TT GAgnAGTATG  550

                      560         570         580         590         600
Hsp 90          551 cTtgGgAGTc TtCTGcTGGA GGTTCCTTCA CTgtGCGTGC TGACcATGGT  600
14201           551 ---------- ---------- ---------- ---------- ----------  600
14201.3         551 .......... .......... .......... .......... ..........  600
14201.5         551 .......... .......... .......... .......... ..........  600
14201.13        551 -TcnGnAGT- TaCTGnTGGA GGTTCCTTCA CTnnGCGTGC TGAC-ATGGT  600

                      610         620         630         640         650
Hsp 90          601 GAGCCCATTG GCAtgGGTAC CAnAGTGATC CTCCATCTCA AAGAAGATCA  650
14201           601 ---------- ---------- ---------- ---------- ----------  650
14201.3         601 .......... .......... .......... .......... ..........  650
14201.5         601 .......... .......... .......... .......... ..........  650
14201.13        601 GAGCCCATnG GgAggGGTAC CAnAGTGATC CTCCATCTCA AAGAAGATCA  650
```

FIGURE 7B

```
                    660        670        680        690        700
Hsp 90      651 GACAGAGTAC CTAGAaGAGA GGCGGgTCAA AGaAGTAGTG AaGAaGCATT 700
14201       651 ---------- ---------- ---------- ---------- ---------- 700
14201.3     651 .......... .......... .......... .......... .......... 700
14201.5     651 .......... .......... .......... .......... .......... 700
14201.13    651 GACAGnGTAC CTAGAnGAGA GGCGGaTCAA AGnAGTAGTG AtGAnGCATc 700

                    710        720        730        740        750
Hsp 90      701 CTCAGtTCAT AGGCTATCCC ATCACCCTTT aTTTGGAGAA GGaACGAGAG 750
14201       701 ---------- ---------- ---------- ---------- ---------- 750
14201.3     701 .......... .......... .......... .......... .......... 750
14201.5     701 .......... .......... .......... .......... .......... 750
14201.13    701 CTCAGaTCAT AGGCTATCCC ATCACCCTTT nTTTGGAGAA GGnACGAGAG 750

                    760        770        780        790        800
Hsp 90      751 AAGGAaATTA GtGATGATGA GGCAGAGGAA GAGAAaGGTG AGAAaGAAGA 800
14201       751 ---------- ---------- ---------- ---------- ---------- 800
14201.3     751 .......... .......... .......... .......... .......... 800
14201.5     751 .......... .......... .......... .......... .......... 800
14201.13    751 AAGGAnATTA GnGATGATGA GGCAGAGGAA GAGAAtGGTG AGAAtGAAGA 800

                    810        820        830        840        850
Hsp 90      801 GGAaGaTAAa GATGATGAAG AAAagCCCAA GATCGAaGAT GTGGgTTCAC 850
14201       801 ---------- ---------- ---------- ---------- ---------- 850
14201.3     801 .......... .......... .......... .......... .......... 850
14201.5     801 .......... .......... .......... .......... .......... 850
14201.13    801 GGAnGnTAAc GATGATGAAG AAAncCCCAA GATCGAtGAT GTGGnTTCAC 850

                    860        870        880        890        900
Hsp 90      851 ATGAGGaGGA TGACAGCGGT aAgGATAAGA AGAAGAAAAC TAaGAagATC 900
14201       851 ---------- ---------- ---------- ---------- ---------- 900
14201.3     851 .......... .......... .......... .......... .......... 900
14201.5     851 .......... .......... .......... .......... .......... 900
14201.13    851 ATGAGGnGGA TGACAGCGGT nAnGATAAGA AGAAGAAnAC TAnGAnnATC 900

                    910        920        930        940        950
Hsp 90      901 AAAGAGAAAT ACATTGATCA GGAAGAACTA AACAAGACCA AGCCTATTTG 950
14201       901 ---------- ---------- ---------- ---------- ---------- 950
14201.3     901 .......... .......... .......... .......... .......... 950
14201.5     901 .......... .......... .......... .......... .......... 950
14201.13    901 .......... .......... .......... .......... .......... 950

                    960        970        980        990       1000
Hsp 90      951 GACCAGAAAC CCTGATGACA TCACCCAAGA GGAGTATGGA GAATTCTACA 1000
14201       951 ---------- ---------- ---------- ---------- ---------- 1000
14201.3     951 .......... .......... .......... .......... .......... 1000
14201.5     951 .......... .......... .......... .......... .......... 1000
14201.13    951 .......... .......... .......... .......... .......... 1000
```

FIGURE 7C

```
                      1010        1020        1030        1040        1050
Hsp 90          1001  AGAGCCTCAC TAATGACTGG GAAGACCACT TGGCAGTCAA GCACTTTTCT  1050
14201           1001  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~  1050
14201.3         1001  .......... .......... .......... .......... ..........  1050
14201.5         1001  .......... .......... .......... .......... ..........  1050
14201.13        1001  .......... .......... .......... .......... ..........  1050

                      1060        1070        1080        1090        1100
Hsp 90          1051  GTAGAAGGTC AGTTGGAATT CAGGGCATTG CTATTTATTC CTCGTCGGGC  1100
14201           1051  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~  1100
14201.3         1051  .......... .......... .......... .......... ..........  1100
14201.5         1051  .......... .......... .......... .......... ..........  1100
14201.13        1051  .......... .......... .......... .......... ..........  1100

                      1110        1120        1130        1140        1150
Hsp 90          1101  TCCCTTTGAC CTTTTTGAGA ACAAGAAGAA AAAGAACAAC ATCAAACTCT  1150
14201           1101  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~AAGAA AAAGAACAAC ATCAAACTCT  1150
14201.3         1101  .......... .......... .......... .......... ..........  1150
14201.5         1101  .......... .......... .......... .......... ..........  1150
14201.13        1101  .......... .......... .......... .......... ..........  1150

                      1160        1170        1180        1190        1200
Hsp 90          1151  ATGTCCGCCG TGTGTTCATC ATGGACAGCT GTGATGAGTT GATACCAGAG  1200
14201           1151  ATGTCCGCCG TGTGTTCATC ATGGnCAGCT GTGATGAGTT GATACCAGAG  1200
14201.3         1151  .......... .......... .......... .......... ..........  1200
14201.5         1151  .......... .......... .......... .......... ..........  1200
14201.13        1151  .......... .......... .......... .......... ..........  1200

                      1210        1220        1230        1240        1250
Hsp 90          1201  TATCTCAATT TTATCCGTGG TGTGGTTGAC TcTGAGGaTC TGCCCCTGAA  1250
14201           1201  TATCTCAATT TTATCCGTGG TGTGGTTGAC TnTGAGGnTC TGCCCCTGAA  1250
14201.3         1201  .......... .......... .......... .......... ..........  1250
14201.5         1201  .......... .......... .......... .......... ..........  1250
14201.13        1201  .......... .......... .......... .......... ..........  1250

                      1260        1270        1280        1290        1300
Hsp 90          1251  CATCTCCCGa GAAATGCTCC AGCAGAGCAA AATCTTGAAA GtCATTCGCA  1300
14201           1251  CATCTCCCGn GAAATGCTCC AGCAGAGCAA AATCTTGAAA GgCATTCGCA  1300
14201.3         1251  .......... .......... .......... .......... ..........  1300
14201.5         1251  .......... .......... .......... .......... ..........  1300
14201.13        1251  .......... .......... .......... .......... ..........  1300

                      1310        1320        1330        1340        1350
Hsp 90          1301  AAAACATTGT TAAGaAGTGC CTTgAGCTCT TCTCTgAGCT GGCAGAAGaC  1350
14201           1301  AAAACATTGT TAAGnAGTGC CTTnAGCTCT TCTCTnAGCT GGCAGAAGnC  1350
14201.3         1301  .......... .......... .......... .......... ..........  1350
14201.5         1301  .......... .......... .......... .......... ..........  1350
14201.13        1301  .......... .......... .......... .......... ..........  1350
```

FIGURE 7D

```
                      1360        1370        1380        1390        1400
Hsp 90          1351 AAGGAGAATT ACAAGAAATT CTATGAGGCA TTCTCTAAAA ATCTCAAGCT 1400
14201           1351 AAGG-GGATT TCAAGAAATT CTTTGGGG-- ---------- ---------- 1400
14201.3         1351 .......... .......... .......... .......... .......... 1400
14201.5         1351 .......... .......... .......... .......... .......... 1400
14201.13        1351 .......... .......... .......... .......... .......... 1400

                      1410        1420        1430        1440        1450
Hsp 90          1401 TGGAATCCAC GAAGACTCCA CTAAGGGCGG CCGGCTGTCT GAGCTGCTGC 1450
14201           1401 ---------- ---------- ---------- ---------- ---------- 1450
14201.3         1401 .......... .......... .......... .......... .......... 1450
14201.5         1401 .......... .......... .......... .......... .......... 1450
14201.13        1401 .......... .......... .......... .......... .......... 1450

                      1460        1470        1480        1490        1500
Hsp 90          1451 GCTATCATAC CTCCCAGTCT GGAGATGAGA TGACATTTCT GTCAGAGTAT 1500
14201           1451 ---------- ---------- ---------- ---------- ---------- 1500
14201.3         1451 .......... .......... .......... .......... .......... 1500
14201.5         1451 .......... .......... .......... .......... .......... 1500
14201.13        1451 .......... .......... .......... .......... .......... 1500

                      1510        1520        1530        1540        1550
Hsp 90          1501 GTTTCTCGCA TGAAGGAGAC ACAGAAGTCC ATCTATTACA TCACTGGTGA 1550
14201           1501 ---------- ---------- ---------- ---------- ---------- 1550
14201.3         1501 .......... .......... .......... .......... .......... 1550
14201.5         1501 .......... .......... .......... .......... .......... 1550
14201.13        1501 .......... .......... .......... .......... .......... 1550

                      1560        1570        1580        1590        1600
Hsp 90          1551 GAGCAAAGAG CAGGTGGCCA ACTCAGCTTT TGTGGAGCGA GTGCGGAAAC 1600
14201           1551 ---------- ---------- ---------- ---------- ---------- 1600
14201.3         1551 .......... .......... .......... .......... .......... 1600
14201.5         1551 .......... .......... .......... .......... .......... 1600
14201.13        1551 .......... .......... .......... .......... .......... 1600

                      1610        1620        1630        1640        1650
Hsp 90          1601 GGGGCTTCGA GGTGGTATAT ATGACCGAGC CCATTGACGA GTACTGTGTG 1650
14201           1601 ---------- ---------- ---------- ---------- ---------- 1650
14201.3         1601 .......... .......... .......... .......... .......... 1650
14201.5         1601 .......... .......... .......... .......... .......... 1650
14201.13        1601 .......... .......... .......... .......... .......... 1650
```

FIGURE 7E

```
                1660        1670        1680        1690        1700
Hsp 90     1651 CAGCAGCTCA AGGAATTTGA TGGGAAGAGC CTGGTCTCAG TTACCAAGGA 1700
14201      1651 ---------- ---------- ---------- ---------- ---------- 1700
14201.3    1651 .......... .......... .......... .......... .......... 1700
14201.5    1651 .......... .......... .......... .......... .......... 1700
14201.13   1651 .......... .......... .......... .......... .......... 1700

                1710        1720        1730        1740        1750
Hsp 90     1701 GGGTCTGGAG CTGCCTGAGG ATGACGAGGA GAAGAAGAAG ATGGAAGAGA 1750
14201      1701 ---------- ---------- ---------- ---.------ .......... 1750
14201.3    1701 .......... .......... .......... .......... .......... 1750
14201.5    1701 .......... .......... .......... .......... .......... 1750
14201.13   1701 .......... .......... .......... .......... .......... 1750

                1760        1770        1780        1790        1800
Hsp 90     1751 GCAAGGCAAA GTTTGAGAAC CTCTGCAAGC TCATGAAAGA AATCTTAGAT 1800
14201      1751 .......... .......... .......... .......... .......... 1800
14201.3    1751 .......... .......... .......... .......... .......... 1800
14201.5    1751 .......... .......... .......... .......... .......... 1800
14201.13   1751 .......... .......... .......... .......... .......... 1800

                1810        1820        1830        1840        1850
Hsp 90     1801 AAGAAGGTTG AGAAGGTGAC AATCTCCAAT AGACTTGTGT CTTCACCTTG 1850
14201      1801 .......... .......... .......... .......... .......... 1850
14201.3    1801 .......... .......... .......... .......... .......... 1850
14201.5    1801 .......... .......... .......... .......... .......... 1850
14201.13   1801 .......... .......... .......... .......... .......... 1850

                1860        1870        1880        1890        1900
Hsp 90     1851 CTGCATTGTG ACCAGCACCT ACGGCTGGAC AGCCAATATG GAGCGGATCA 1900
14201      1851 .......... .......... .......... .......... .......... 1900
14201.3    1851 .......... .......... .......... .......... .......... 1900
14201.5    1851 .......... .......... .......... .......... .......... 1900
14201.13   1851 .......... .......... .......... .......... .......... 1900

                1910        1920        1930        1940        1950
Hsp 90     1901 TGAAAGCCCA GGCACTTCGG GACAACTCCA CCATGGGCTA TATGATGGCC 1950
14201      1901 .......... .......... .......... .......... .......... 1950
14201.3    1901 .......... .......... .......... .......... .......... 1950
14201.5    1901 .......... .......... .......... .......... .......... 1950
14201.13   1901 .......... .......... .......... .......... .......... 1950

                1960        1970        1980        1990        2000
Hsp 90     1951 AAAAAGCACC TGGAGATCAA CCCTGACCAC CCCATTGTGG AGACGCTGCG 2000
14201      1951 .......... .......... .......... .......... .......... 2000
14201.3    1951 .......... .......... .......... .......... .......... 2000
14201.5    1951 .......... .......... .......... .......... .......... 2000
14201.13   1951 .......... .......... .......... .......... .......... 2000
```

FIGURE 7F

```
                    2010        2020        2030        2040        2050
Hsp 90      2001 GCAGAAGGCT GAGGCCGACA AGAATGATAA GGCAGTTAAG GACCTGGTGG 2050
14201       2001 .......... .......... .......... .......... .......... 2050
14201.3     2001 .......... .......... .......... .......... .......... 2050
14201.5     2001 .......... .......... .......... .......... .......... 2050
14201.13    2001 .......... .......... .......... .......... .......... 2050

                    2060        2070        2080        2090        2100
Hsp 90      2051 TGCTGCTGTT TGAAACDGCC CTGCTATCTT CTGGCTTTTC CCTTGAGGAT 2100
14201       2051 .......... .......... .......... .......... .......... 2100
14201.3     2051 .......... .......... .......... .......... .......... 2100
14201.5     2051 .......... .......... .......... .......... .......... 2100
14201.13    2051 .......... .......... .......... .......... .......... 2100

                    2110        2120        2130        2140        2150
Hsp 90      2101 CCCCAGACCC ACTCCAACCG CATCTATCGC ATGATCAAGC TAGGTCTAGG 2150
14201       2101 .......... .......... .......... .......... .......... 2150
14201.3     2101 .......... .......... .......... .......... .......... 2150
14201.5     2101 .......... .......... .......... .......... .......... 2150
14201.13    2101 .......... .......... .......... .......... .......... 2150

                    2160        2170        2180        2190        2200
Hsp 90      2151 TATTGATGAA GATGAAGTGG CAGCAGNGGA ACCCAATGCT GCAGTTCCTG 2200
14201       2151 .......... .......... .......... .......... .......... 2200
14201.3     2151 .......... .......... .......... .......... .......... 2200
14201.5     2151 .......... .......... .......... .......... .......... 2200
14201.13    2151 .......... .......... .......... .......... .......... 2200

                    2210        2220        2230        2240        2250
Hsp 90      2201 ATGAGATCCC CCCTCTCGAC GGCGATGAGG ATGCGTCTCG CATGGAAGAA 2250
14201       2201 .......... .......... .......... .......... .......... 2250
14201.3     2201 .......... .......... .......... .......... .......... 2250
14201.5     2201 .......... .......... .......... .......... .......... 2250
14201.13    2201 .......... .......... .......... .......... .......... 2250

                    2260        2270        2280        2290        2300
Hsp 90      2251 GTCGATTAGG TTAGGAGTTC ATAGTTGGAA AACTTGTGCC CTTGTAAAGT 2300
14201       2251 .......... .......... .......... .......... .......... 2300
14201.3     2251 .......... .......... .......... .......... .......... 2300
14201.5     2251 .......... .......... .......... .......... .......... 2300
14201.13    2251 .......... .......... .......... .......... .......... 2300

                    2310        2320        2330        2340        2350
Hsp 90      2301 GTCCCCATGG GCTCCCACTG CAGCCTCGAG TGCCCCTGTC CCACCTGGCT 2350
14201       2301 .......... .......... .......... .......... .......... 2350
14201.3     2301 .......... .......... .......... .......... .......... 2350
14201.5     2301 .......... .......... .......... .......... .......... 2350
14201.13    2301 .......... .......... .......... .......... .......... 2350
```

FIGURE 7G

```
                    2360        2370        2380        2390        2400
Msp 90      2351 CCCCCTGCTG GTGTCTAGTG TTTTTTTCCC TCTCCTGTCC TTGTGTTGAA 2400
14201       2351 .......... .......... .......... .......... .......... 2400
14201.3     2351 .......... .......... .......... .......... .......... 2400
14201.5     2351 .......... .......... .......... .......... .......... 2400
14201.13    2351 .......... .......... .......... .......... .......... 2400

                    2410        2420        2430        2440        2450
Msp 90      2401 GGCAGTAAAC TAAGGGTGTC AAGCCCCATT CCCTCTCTAC TCTTGACAGC 2450
14201       2401 .......... .......... .......... .......... .......... 2450
14201.3     2401 .......... .......... .......... .......... .......... 2450
14201.5     2401 .......... .......... .......... .......... .......... 2450
14201.13    2401 .......... .......... .......... .......... .......... 2450

                    2460        2470        2480        2490        2500
Msp 90      2451 AGGATTGGAT GTTGTGTATT GTGGTTTATT TTATTTTCTT CATTTTGTTC 2500
14201       2451 .......... .......... .......... .......... .......... 2500
14201.3     2451 .......... .......... .......... .......... .......... 2500
14201.5     2451 .......... .......... .......... .......... .......... 2500
14201.13    2451 .......... .......... .......... .......... .......... 2500

                    2510        2520        2530        2540        2550
Msp 90      2501 TGAAATTAAA GTATGCAAAA TAAAGAATAT GCCGTTTTTA TAC....... 2550
14201       2501 .......... .......... .......... .......... .......... 2550
14201.3     2501 .......... .......... .......... .......... .......... 2550
14201.5     2501 .......... .......... .......... .......... .......... 2550
14201.13    2501 .......... .......... .......... .......... .......... 2550
```

FIGURE 7H

```
                      10          20          30          40          50
capthepsin          1 TCCGGCAACG CCAACCGCTC CGCTGCGCGC AGGCTGGGCT GCAGGCTCTC  50
87058               1 ---------- ---------- ---------- ---------- ----------  50
87058.6             1 ---------- ---------- ---------- ---------- ----------  50
87058.8             1 ---------- ---------- ---------- ---------- ----------  50
87058.16            1 ---------- ---------- ---------- ---------- ----------  50

                      60          70          80          90         100
capthepsin         51 GGCTGCAGCG CTGGGCTGGT GTGCAGTGGT GCGACCACGG CTCACGGCAG 100
87058              51 ---------- ---------- ---------- ---------- ---------- 100
87058.6            51 ---------- ---------- ---------- ---------- ---------- 100
87058.8            51 ---------- ---------- ---------- ---------- ---------- 100
87058.16           51 ~~~~~~~~~NCN GGTTGAGNAT TCGGACNAGT CCGAAAACGT CCGGCAAGTC 100

                     110         120         130         140         150
capthepsin        101 CCTCAGCCAC CCAGATGTAA GCGATCTGGT TCCCACCTCA GCCTCCGGAG 150
87058             101 ---------- ---------- ---------- ---------- ---------- 150
87058.6           101 ---------- ---------- ---------- ---------- ---------- 150
87058.8           101 ---------- ---------- ---------- ---------- ---------- 150
87058.16          101 ACCCGCTCCG CTGNGCGCAG GCTGGGNTGC AGGCTCTCGG NTGCAGNGCT 150

                     160         170         180         190         200
capthepsin        151 TAGTGGATCT AGGATCCGGC TTCCAACATG TGGCAGcTCT GGGCCTCCCT 200
87058             151 ---------- ---------- ---------- ---------- ---------- 200
87058.6           151 ---------- ---------- ---------- ---------- ---------- 200
87058.8           151 ---------- ---------- ---------- ---------- ---------- 200
87058.16          151 GGGTGGATCT AGGATCCGGC TTCCAACATG TGGCAGTTCT GGGCCTCCCT 200

                     210         220         230         240         250
capthepsin        201 CTGcTGCCTG CTGGTGTTGG cCAATGCCCG GAGcAGGcCC TCTTTCCATC 250
87058             201 ---------- ---------- ---------- ---------- ---------- 250
87058.6           201 ---------- ---------- ---------- ---------- ---------- 250
87058.8           201 ---------- ---------- ---------- ---------- ---------- 250
87058.16          201 CTGnTGCCTG CTGGTGTTGG aCAATGCCCG GAGgAGGnCC TCTTTCCATC 250

                     260         270         280         290         300
capthepsin        251 CCCTGTCGGA TGAGCTGGTC AaCTATGTCA ACAAACGGAA TACCACGTGG 300
87058             251 ---------- ---------- ---------- ---------- ---------- 300
87058.6           251 ---------- ---------- ---------- ---------- ---------- 300
87058.8           251 ---------- ---------- ---------- ---------- ---------- 300
87058.16          251 CCCTGTCGGA TGAGCTGGTC AnCTATGTCA ACAAACGGAA TACCACGTGG 300
```

FIGURE 8A

```
                    310         320         330         340         350
capthepsin      301 cAGGCCGGaA ACAACTTCTA CAACGTGGAC ATGAGCTACT TGAaGAGGcT 350
87058           301 ---------- ---------- ---------- ---------- ---------- 350
87058.6         301 ---------- ---------- ---------- ---------- ---------- 350
87058.8         301 ---------- ---------- ---------- ---------- ---------- 350
87058.16        301 nAGGCCGGgA ACAACTTCTA CAACGTGGAC ATGAGCTACT TGAnGAGGnT 350

                    360         370         380         390         400
capthepsin      351 ATGTGGTACC TTCCTGGGTG GGCCCAAGCC ACCCCAGAGA GTTATGTTTA 400
87058           351 ---------- ---------- ---------- ---------- ---------- 400
87058.6         351 ---------- ---------- ---------- ---------- ---------- 400
87058.8         351 --GaGGTACC TTCCTGGGTG GGCCCAAGCC ACCCCAGAGA GTTATGTTTA 400
87058.16        351 ATGTGGTACC TTCCTGGGTG GGCCCAAGCC ACCCCAGAGA GTTNTGTTTA 400

                    410         420         430         440         450
capthepsin      401 CCGAGGACCT GAAGCTGCCT GCAAGCTTCG ATCCACGGGA ACAATGGCCA 450
87058           401 ---------- ---------- ---------- ---------- ---------- 450
87058.6         401 ---------- ---------- ---------- ---------- ---------- 450
87058.8         401 CCGAGGACCT GAAGCTGCCT GCAAGCTTCG ATGCACGGGA ACAATGGCCA 450
87058.16        401 CCGAGGACCT GAAGCTGCCT GCAAGCTTCG AaGgACGGGA ACAATGGCCA 450

                    460         470         480         490         500
capthepsin      451 CAGTGTCCCA CCATCAAAGA GATCAGAGAC CAGGGCTCCT GTGGCTCCTG 500
87058           451 ---------- ---------- ---------- ---------- ---------- 500
87058.6         451 ---------- ---------- ---------- ---------- ---------- 500
87058.8         451 CACTGTCCCA CCATCAAAGA GATCAGAGAC CAGGGNTCCT GTGGCTCCTG 500
87058.16        451 CAGTGTCCCA CCATCAAAGA GATCAGAGAN CAGGGCTCCT GTGGNTCCTG 500

                    510         520         530         540         550
capthepsin      501 CTGGGCCTTC GGGGCTGTGG AAGCCATCTC TGACCGGATC TGCATCCACA 550
87058           501 ---------- ---------- ---------- ---------- ---------- 550
87058.6         501 ---------- ---------- ---------- ---------- ---------- 550
87058.8         501 CTGGGCCTTC GGGGCTGTGG AAGCCATCTC TGACCGGATC TGNATCCACA 550
87058.16        501 CTGGGCCTeC GGGGCTGTGG AAGNCATCTC TGACCGGATC TGCATCCACA 550

                    560         570         580         590         600
capthepsin      551 CCAATGGGCA CGTCAGCGTG GAGGTGTCGG GGGAGGACCT GCTCACATGC 600
87058           551 ---------- ---------- ---------- ---------- ---------- 600
87058.6         551 ---------- ---------- ---------- ---------- ---------- 600
87058.8         551 CCAATGCGCA CGTCAGCGTG GAGGTGTCGG CGGAGGAC-T GCTCACATGC 600
87058.16        551 CCAATGNGCA CGTCAGCGTG GtGGTGTCGG NGGAGGACCT GaTCACCTNt 600

                    610         620         630         640         650
capthepsin      601 TGTGGCAGCA TGTGTGGGGA CGGCTGTAAT GGTGGCTATC CTGCTGAAGC 650
87058           601 ---------- ---------- ---------- ---------- ---------- 650
87058.6         601 ---------- ---------- ---------- ---------- ---gTGAAGC 650
87058.8         601 TGTGGCAGNA TGTGTGGGGA CGGCTGTAAT GGTGGCTATC CTGCTGAAGC 650
87058.16        601 TGTGGtAGCA TGTGTGGGGA CGGCTGTAAT GGtGGCTATC CTGNTGAAGC 650
```

FIGURE 8B

```
                    660        670        680        690        700
capthepsin      651 TTGGAACTTC TGGACAAGAA AAGGCCTGGT TTCTGGTGCC CTCTATGAAT 700
87058           651 ---------- ---------- ---------- ---------- ---------- 700
87058.6         651 TTGGAACTTC TGGACAAGAA AAGGCCTGGT TTCTGGTGGC CTCTATGAAT 700
87058.8         651 TTGGNACTTC TGGACAAGAA AAGGCCTGGT TTCTGGTGGC CTCTATGANT 700
87058.16        651 TNGGgNCTTC TNagaAAGAA AAGGCtNGtT TT--GGTGGC CT-TATGAcT 700

                    710        720        730        740        750
capthepsin      701 CCCATGTAGG GTGCAGACCG TACTCCATCC CTCCCTGTGA GCACCACGTC 750
87058           701 ---------- ---------- ---------- ---------- ---------- 750
87058.6         701 CCCATGTAGG GTGCAGACCG TACTCCATCC CTCCCTGTGA GCACCACGTC 750
87058.8         701 CCCATGTAGG GTGTAGACCG TACTTCATCC CTCCCTGTGA GCACCACGTC 750
87058.16        701 CCCATGT... ..........  .......... .......... .......... 750

                    760        770        780        790        800
capthepsin      751 AACGGCTCCC GGCCCCCATG CACGGGGGAG GGAGATACCC CCAAGTGTAG 800
87058           751 ---------- ---------- ---------- ---------- ---------- 800
87058.6         751 AACGGCTCCC GGCCCCCATG CACGGGGGAG GGAGATACCC CCAAGTGTAG 800
87058.8         751 AACGGGLTCCC GGgCCCCATG CACGGNGGAG GGAGATACCC CCAAGTGTAa 800
87058.16        751 .......... .......... .......... .......... .......... 800

                    810        820        830        840        850
capthepsin      801 CAAGATCTGT GAGCCTGGCT ACAGCCCGAC CTACAAACAG GACAAGCACT 850
87058           801 ---------- ---------- ---------- ---------- ---------- 850
87058.6         801 CAAGATCTGT GAGCCTGGCT ACAGCCCGAC CTACAAACAG GACAAGCACT 850
87058.8         801 CAAGATCTGT GAGCCTGGgT ACAGtCCcga CtACAAACAG GAaAAGCACT 850
87058.16        801 .......... .......... .......... .......... .......... 850

                    860        870        880        890        900
capthepsin      851 ACGGATACAA TTCCTACAGC GTCTCCAATA GCGAGAAGGA CATCATGGCC 900
87058           851 ---------- ---------- ---------- ---------- ---------- 900
87058.6         851 ACGGATACAA TTCCTACAGC GTCTCCAATA GCGAGAAGGA CATCATGGCC 900
87058.8         851 ACGGATACAA TTCCT-CAGN GTCTCCAATA GtGAGAAGGA CATCAT-GCC 900
87058.16        851 .......... .......... .......... .......... .......... 900

                    910        920        930        940        950
capthepsin      901 GAGATCTACA AAAACGGCCC CGTGGAGGGA GCTTTCTCTG TGTATTCGGA 950
87058           901 ---------- ---------- ---------- ---------- ---------- 950
87058.6         901 GAGATCTACA AAAACGGCCC CGTGGAGGGA GCTTTCTCTG TGTATTCGGA 950
87058.8         901 GAGATCTACA AtAACGGC.. .......... .......... .......... 950
87058.16        901 .......... .......... .......... .......... .......... 950

                    960        970        980        990       1000
capthepsin      951 CTTCCTGCTC TACAAGTCAG GAGTGTACCA ACACGTCACC GGAGAGATGA 1000
87058           951 ---------- ---------- ---------- ---------- ---------- 1000
87058.6         951 CTTCCTGCTC TACAAGTCAG GAGTGTACCA ACACGTCACC GGAGAGATGA 1000
87058.8         951 .......... .......... .......... .......... .......... 1000
87058.16        951 .......... .......... .......... .......... .......... 1000
```

FIGURE 8C

```
                    1010        1020        1030        1040        1050
capthepsin      1001 TGGGTGGCCA TGCCATCCGC ATCCTGGGCT GGGGAGTGGA GAATGGCACA 1050
87058           1001 ---------- ---------- ---------- ---------- ---------- 1050
87058.6         1001 TGGGTGGCCA TGCCATCCGC ATCCTGGGCT GGGGAGTGGA GAATGGCACA 1050
87058.8         1001 .......... .......... .......... .......... .......... 1050
87058.16        1001 .......... .......... .......... .......... .......... 1050

                    1060        1070        1080        1090        1100
capthepsin      1051 GCCTACTGGC TGGTTGCCAA CTCCTGGAAC ACTGACTGGG GTGACAATGG 1100
87058           1051 ~~~~~~~~cGg cagacGCCAA CTCCTGGAAC ACTGACTGGG GTGACAATGG 1100
87058.6         1051 aCCTACTGGC TGGTTGqCAA CTCCTGGAAC ACTGACTGGG GTGACAATGG 1100
87058.8         1051 .......... .......... .......... .......... .......... 1100
87058.16        1051 .......... .......... .......... .......... .......... 1100

                    1110        1120        1130        1140        1150
capthepsin      1101 CTTCTTTAAA ATACTCAGAG GACAGGATCA CTGTGGAATC GAATCAGAAG 1150
87058           1101 CTTCTTTAAA ATACTCAGAG GACAGGTTCA CTGTGGAATC GAATCAGAAG 1150
87058.6         1101 gTTC~~~~~~ ---------- ---------- ---------- ---------- 1150
87058.8         1101 .......... .......... .......... .......... .......... 1150
87058.16        1101 .......... .......... .......... .......... .......... 1150

                    1160        1170        1180        1190        1200
capthepsin      1151 TGGTGGCTGG AATTCCACGC ACCGATCAGT ACTGGGAAAA GATCTAATCT 1200
87058           1151 TGGTGGCTGG AATTCCACGC ACCGTTCAGT ACTGGGAAAA GNTCTAATCT 1200
87058.6         1151 ---------- ---------- ---------- ---------- ---------- 1200
87058.8         1151 .......... .......... .......... .......... .......... 1200
87058.16        1151 .......... .......... .......... .......... .......... 1200

                    1210        1220        1230        1240        1250
capthepsin      1201 GCCGTGGGCC TGTCGTGCCA GTCCTGGGGG CGAGATCGGG GTAGAAATGC 1250
87058           1201 GCCGTGGGCC TNTCGTGCCA GTCCTGGGGG CGAGATGGGG GTAGAAATGC 1250
87058.6         1201 ---------- ---------- ---------- ---------- ---------- 1250
87058.8         1201 .......... .......... .......... .......... .......... 1250
87058.16        1201 .......... .......... .......... .......... .......... 1250

                    1260        1270        1280        1290        1300
capthepsin      1251 ATTTTATTCT TTAAGTTCAC GTAAGATACA AGTTTCAGgC AGGGTCTGAA 1300
87058           1251 ATTTTATTCT TTAAGTTCAC GTAAGATACA AGTTTCAGaC AGGGTCTnAA 1300
87058.6         1251 .......... .......... .......... .......... .......... 1300
87058.8         1251 .......... .......... .......... .......... .......... 1300
87058.16        1251 .......... .......... .......... .......... .......... 1300

                    1310        1320        1330        1340        1350
capthepsin      1301 GGACTGGaTT gGCCAAACAT CAGACCTGTC TTCCAAGGAG ACCAAGTCCT 1350
87058           1301 GGaCTGGcTT nGCCAAAnAT CAGACCTGT. .......... .......... 1350
87058.6         1301 .......... .......... .......... .......... .......... 1350
87058.8         1301 .......... .......... .......... .......... .......... 1350
87058.16        1301 .......... .......... .......... .......... .......... 1350
```

FIGURE 8D

```
                1360        1370        1380        1390        1400
cepthepsin   1351 GGCTACATCC CAGCCTGTGG TTACAGTGCA GACAGGCCAT GTGAGCCACC 1400
87058        1351 .......... .......... .......... .......... .......... 1400
87058.6      1351 .......... .......... .......... .......... .......... 1400
87058.8      1351 .......... .......... .......... .......... .......... 1400
87058.16     1351 .......... .......... .......... .......... .......... 1400

                1410        1420        1430        1440        1450
cepthepsin   1401 GCTGCCAGCA CAGAGCGTCC TTCCCCCTGT AGACTAGTGC CGTGGGAGTA 1450
87058        1401 .......... .......... .......... .......... .......... 1450
87058.6      1401 .......... .......... .......... .......... .......... 1450
87058.8      1401 .......... .......... .......... .......... .......... 1450
87058.16     1401 .......... .......... .......... .......... .......... 1450

                1460        1470        1480        1490        1500
cepthepsin   1451 CCTGCTGCCC AGCTGCTGTG GCCCCCTCCG TGATCCATCC ATCTCCAGGG 1500
87058        1451 .......... .......... .......... .......... .......... 1500
87058.6      1451 .......... .......... .......... .......... .......... 1500
87058.8      1451 .......... .......... .......... .......... .......... 1500
87058.16     1451 .......... .......... .......... .......... .......... 1500

                1510        1520        1530        1540        1550
cepthepsin   1501 AGCAAGACAG AGACGCAGGA TGGAAAGCGG AGTTCCTAAC AGGATGAAAG 1550
87058        1501 .......... .......... .......... .......... .......... 1550
87058.6      1501 .......... .......... .......... .......... .......... 1550
87058.8      1501 .......... .......... .......... .......... .......... 1550
87058.16     1501 .......... .......... .......... .......... .......... 1550

                1560        1570        1580        1590        1600
cepthepsin   1551 TTCCCCCATC AGTTCCCCCA GTACCTCCAA GCAAGTAGCT TTCCACATTT 1600
87058        1551 .......... .......... .......... .......... .......... 1600
87058.6      1551 .......... .......... .......... .......... .......... 1600
87058.8      1551 .......... .......... .......... .......... .......... 1600
87058.16     1551 .......... .......... .......... .......... .......... 1600

                1610        1620        1630        1640        1650
cepthepsin   1601 GTCACAGAAA TCAGAGGAGA GATGGTGTTG GGAGCCCTTT GGAGAACGCC 1650
87058        1601 .......... .......... .......... .......... .......... 1650
87058.6      1601 .......... .......... .......... .......... .......... 1650
87058.8      1601 .......... .......... .......... .......... .......... 1650
87058.16     1601 .......... .......... .......... .......... .......... 1650
```

FIGURE 8E

```
                      1660        1670        1680        1690        1700
capthepsin      1651 AGTCTCCAGG TCCCCCTGCA TCTATCGAGT TTGCAATGTC ACAACCCTCC 1700
87058           1651 .......... .......... .......... .......... .......... 1700
87058.6         1651 .......... .......... .......... .......... .......... 1700
87058.8         1651 .......... .......... .......... .......... .......... 1700
87058.16        1651 .......... .......... .......... .......... .......... 1700

                      1710        1720        1730        1740        1750
capthepsin      1701 TGATCTTGTG CTCAGCATGA TTCTTTAATA GAAGTTTTAT TTTTCGTGCA 1750
87058           1701 .......... .......... .......... .......... .......... 1750
87058.6         1701 .......... .......... .......... .......... .......... 1750
87058.8         1701 .......... .......... .......... .......... .......... 1750
87058.16        1701 .......... .......... .......... .......... .......... 1750

                      1760        1770        1780        1790        1800
capthepsin      1751 CTCTGCTAAT CATGTGGGTG AGCCAGTGGA ACAGCGGGAG CCTGTGCTGG 1800
87058           1751 .......... .......... .......... .......... .......... 1800
87058.6         1751 .......... .......... .......... .......... .......... 1800
87058.8         1751 .......... .......... .......... .......... .......... 1800
87058.16        1751 .......... .......... .......... .......... .......... 1800

                      1810        1820        1830        1840        1850
capthepsin      1801 TTTGCAGATT GCCTCCTAAT GACGCGGCTC AAAAGGAAAC CAAGTGGTCA 1850
87058           1801 .......... .......... .......... .......... .......... 1850
87058.6         1801 .......... .......... .......... .......... .......... 1850
87058.8         1801 .......... .......... .......... .......... .......... 1850
87058.16        1801 .......... .......... .......... .......... .......... 1850

                      1860        1870        1880        1890        1900
capthepsin      1851 GGAGTTGTTT CTGACCCACT GATCTCTACT ACCACAAGGA AAATAGTTTA 1900
87058           1851 .......... .......... .......... .......... .......... 1900
87058.6         1851 .......... .......... .......... .......... .......... 1900
87058.8         1851 .......... .......... .......... .......... .......... 1900
87058.16        1851 .......... .......... .......... .......... .......... 1900

                      1910        1920        1930        1940        1950
capthepsin      1901 GGAGAAACCA GCTTTTACTG TTTTTGAAAA ATTACAGCTT CACCCTGTCA 1950
87058           1901 .......... .......... .......... .......... .......... 1950
87058.6         1901 .......... .......... .......... .......... .......... 1950
87058.8         1901 .......... .......... .......... .......... .......... 1950
87058.16        1901 .......... .......... .......... .......... .......... 1950

                      1960        1970        1980        1990        2000
capthepsin      1951 AGTTAACAAG GAATGCCTGT GCCAATAAAA GGTTTCTCCA ACTTGA.... 2000
87058           1951 .......... .......... .......... .......... .......... 2000
87058.6         1951 .......... .......... .......... .......... .......... 2000
87058.8         1951 .......... .......... .......... .......... .......... 2000
87058.16        1951 .......... .......... .......... .......... .......... 2000
```

FIGURE 8F

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/08501

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68 C12P19/34 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS. EDITOR INNIS M.; PUBLISHER ACADEMIC, 1990, SAN DIEGO, CALIF., pages 219-27, XP002015609 OCHMAN, H. ET AL: "Amplification of flanking sequences by inverse PCR" see whole article ---	1-8
X	BIOTECHNIQUES, vol. 18, no. 5, May 1995, pages 762-64, XP000509322 COOLIDGE C ET AL: "Run-around PCR: A novel way to create duplications using polymerase chain reaction " see the whole document --- -/--	1-8

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 October 1996	25.10.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Osborne, H

Form PCT/ISA/210 (second sheet) (July 1992)

C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JOURNAL OF BIOLOGICAL CHEMISTRY., vol. 268, no. 12, 1993, pages 8842-50, XP000604943 LEE, D. ET AL.: "Molecular cloning and genomic organization of a gene for luciferin-binding protein from dinoflagellate Gonyaulax polyedra" see the whole document ---	1-8
X	US,A,4 994 370 (SILVER) 19 February 1991 see the whole document ---	1-8
X	JOURNAL OF VIROLOGICAL METHODS, vol. 49, no. 3, January 1994, pages 269-84, XP000606337 TSUEI D-J ET AL: "Inverse polymerase chain reaction for cloning cellular sequences adjacent to integrated hepatitis b virus in hepatocellular carcinomasw" see the whole document ---	1-8
X	WO,A,90 14423 (THE INFERGENE CO.) 29 November 1990 see page 19 ---	1-8
A	WO,A,93 12257 (HYBRITECH INC) 24 June 1993 see the whole document ---	1-8
A	NUCLEIC ACIDS RESEARCH, vol. 19, 1991, pages 3055-60, XP002015610 PARKER J. ET AL: "Walking PCR" cited in the application -----	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PC./US 96/08501

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US-A-4994378	19-02-91	NONE	
WO-A-9014423	29-11-90	NONE	
WO-A-9312257	24-06-93	AU-A- 3274793	19-07-93
		US-A- 5512463	30-04-96